9-9-2020 1:30 PM

Analysis of Secondary Metabolites Biosynthesized by Pathogenic and Symbiotic Fungi using High-Resolution Tandem LC-MS and Spectral Molecular Networking

Natasha DesRochers, *The University of Western Ontario*

Supervisor: Sumarah, Mark W., *The University of Western Ontario*
Co-Supervisor: Yeung, Ken K.-C., *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Chemistry

Thesis abstract

Fungi from Canadian crops affect many aspects of agriculture. For example, ginseng root rot caused by *Ilyonectria* spp. is a major issue for farmers. *Ilyonectria* spp. make few reported natural products and it is not known if those products are unique to virulent species. PCA and molecular networking were applied to HRMS data to establish a distinct metabolomic profile of root rot pathogens, characterized by antifungal resorcyclic acid lactones (RALs). These likely protect *Ilyonectria* from other soil pathogens, which gives it the opportunity to infect ginseng root.

Molecular networking was also applied to fungal endophytes from fruit crops to identify new compounds. Endophytes are symbionts that produce beneficial compounds, however, traditional screening methods are cumbersome in finding new or novel compounds. HRMS data of 302 fungal endophytes were examined to identify nine new compounds related to known antimicrobial compounds. These compounds are targets for isolating and characterizing in further studies.

Keywords: metabolomics, ginseng root rot, *Ilyonectria*, endophytes, untargeted analysis, mass spectrometry, secondary metabolites, natural products, molecular networking, GNPS

Summary for Lay Audience

There are many kinds of fungi that affect agriculture, ranging from harmful to helpful. Often the effects they have on plants are due to chemicals they produce, which are named natural products. Ginseng root rot is caused by a fungal infection and results in large crop losses for Canadian farmers every year. The pathogen that causes this disease was studied to determine what kinds of natural products it makes, and what role those products may have in the life of the fungus and its host, ginseng. Mass spectrometry is an analytical technique that measures the mass of chemicals and also allows us to break chemicals into fragments. Similarly-shaped chemicals produce similar types of fragments and this similarity can be assessed by a technique named molecular networking. Using molecular networking it was determined that fungi that cause ginseng root rot produce a family of structurally related natural products that can kill other types of fungi. These compounds likely help ginseng root rot fungus ward off other fungi living in its environment to give it a better chance of survival, which in turn allows it to infect ginseng root.

Molecular networking was also applied to another group of fungi to identify new compounds. Fungal endophytes grow in mutually beneficial relationships with plants. They grow inside leaves and stems without harming their host and produce natural products that protect their hosts from other more harmful organisms. In return they receive a safe place to live and nutrients to grow. Many of the chemicals they make are used as medicinal drugs, pesticides, or have value in chemical research. By examining fungi and their natural products we can discover new chemicals that may be of use. However, the processes used to find new compounds from fungi are often lengthy and difficult. Therefore, 302 fungal endophytes found in Canadian fruit crops were analyzed by molecular networking to identify undiscovered natural products. Molecular networking simplified and sped up the process of finding new compounds. Nine new compounds

were found that are related to known compounds that have antimicrobial activity. These compounds will be isolated and have their structures characterized in future studies.

Co-authorship Statement

Chapter 2 – Metabolomic Profiling of Ginseng Root Rot Fungi

Experimental design, experiments, data analyses, and manuscript preparation were performed by the author. Seed spectra compounds for molecular networking were isolated by Jacob Walsh (PhD Candidate, Western University-Chemistry Department). Justin Renaud (PhD, LORDC) assisted with methodology and compound identification. Mark Sumarah (PhD, LORDC) and Ken K.-C. Yeung (PhD, Western University-Biochemistry and Chemistry Departments) supervised. Fungal cultures were provided by Keith Seifert (PhD, AAFC Ottawa) from the Canadian Collection of Fungal Cultures (Ottawa, Ontario), who also gave detailed and current information on fungal taxonomy. All authors contributed to reviewing and editing. Grateful acknowledgments are also given to Megan Kelman (LORDC) and Amy McMillan (PhD, Cleveland Clinic) for technical support with R.

Acknowledgments

Firstly, thank you to my friends and family for all the love and help you have given me over the past two years. I am lucky to have such a strong community supporting me, and I appreciate you all so much.

A thousand thanks to my colleagues Megan Kelman and Lyne Sabourin. Your endless patience and kindness have made every day of my studies enjoyable rather than stressful. Megan, thank you for all the assistance with R-related nonsense and for all the great life pro-tips you have given me. Lyne, you have been an amazing cubicle-mate. I have had so much fun organizing lab activities and attempting to grow a little windowsill garden with you. Thank you both for teaching me how to calibrate the mass spec so many times!

Thank you to Dr. Justin Renaud for teaching me most of what I know about untargeted mass spec analysis. You have been so encouraging of my research career since I first started at AAFC as an intern more than four years ago. Your constant enthusiasm for science makes me excited to learn more and to do better as a researcher. Thank you as well to my good friend Jacob Walsh. We have so much fun in the lab whether it is designing new experiments, doing sample prep, or washing dishes. Thanks for being such a supportive friend and lab-mate.

Thank you to the many interns, graduate students, and technicians at the London Research and Development Centre who have helped me since I was an intern. An especially big thank you to interns Wayne Allison, Jake Artibello, Friday Black, Lucy Kim, Jared King, Janet Schapurga, Ben Topp, and Anthony Truong; students Steph Collins, Juan Li, Cameron Littlejohn, Amy McMillan, Luc Morrison and Katherine Teeter; technicians Shawn Hoogstra, Emine Kapanoglu, Sophie Krolikowski, and Tim McDowell; and visiting professors Jian Zhang and Yanshen Li.

Thank you as well to LORDC graphics expert Alex Molnar for your help with conference posters, graphical abstracts, and experimental photos.

Thank you to Dr. Sean Westerveld and Carl Atkins at OMAFRA for their expertise on ginseng root rot and agricultural practices. I also immensely appreciate the wisdom of Dr. Keith Seifert and Dr. J. D. Miller. Thank you for lending your perspectives as long-time experts in mycology and natural products chemistry. Thank you to Dr. Ashraf Ibrahim, Dr. David McMullin, and Dr. Joey Tanney for sharing your knowledge on fungal endophytes and their secondary metabolites. Thank you as well to my committee, Dr. Lars Konermann, Dr. Martin Stillman, and Dr. Brad Urquhart.

Thank you to my co-supervisor, Dr. Ken Yeung. Your assistance allowed me to attend the Gordon Research Conference on Mycotoxins and Phycotoxins, at which I gained critical insight into solving my longstanding molecular networking troubles. This event truly changed the course of my thesis for the better. Throughout my degree you have been incredibly supportive and encouraging, for which I am grateful. Thank you also to Chaochao Chen and Kristina Jurcic from Dr. Yeung's lab for taking time to teach me about MALDI.

Finally, I extend all my gratitude to my supervisor, Dr. Mark Sumarah. Thank you for hiring me as an intern four years ago, and for taking me on as a graduate student. You have been the best supervisor I could have asked for and you have really made my degree so fulfilling. Thank you for giving me space to work independently but also for pushing me to do more. I am very proud of everything I've done in the past four years but wouldn't be nearly as accomplished as I am without your help.

Contents

List of Tables

List of Figures

List of Symbols and Abbreviations

AAFC: Agriculture and Agri-Food Canada
AC: alternating current
AGC: automatic gain control
amu: atomic mass unit
API: atmospheric pressure ionization
BH: Benjamini-Hochberg
BLAST: basic local alignment search tool
CCFC: Canadian Collection of Fungal Cultures
CID: collision-induced dissociation
CRM: charge residue model
Da: Daltons
DAOMC: Department of Agriculture, Ottawa, Mycological Collection (historical acronym for CCFC, still in use)
DDA: data-dependent acquisition
DESI: desorption electrospray ionization
DFF: diagnostic fragmentation filtering
DNA: deoxyribonucleic acid
DON: deoxynivalenol
EIC: extracted ion chromatogram
ESI: electrospray ionization
FT: Fourier transform
FT-ICR: Fourier transform ion cyclotron resonance
FWHM: full width at half maximum
GNPS: Global natural products social molecular networking
HCD: Higher energy collisional dissociation
HESI: heated electrospray ionization
HPLC: high-performance liquid chromatography
HRMS: high-resolution mass spectrometry
HSP90: heat shock protein 90

IEM: ion evaporation model

ITS: internal transcribed spacer

KW: Kruskal Wallis

kV: kilovolts

LC: liquid chromatography

LIT: linear ion trap

LORDC: London Research and Development Centre

m/z: mass-to-charge ratio

MALDI: matrix-assisted laser desorption ionization

Max IT: maximum injection time

METLIN: METabolite LINk (online tandem mass spectra repository)

MS: mass spectrometry

MS/MS: tandem mass spectrometry

NCE: normalized collision energy

NLS: neutral loss scanning

NMR: nuclear magnetic resonance

PCA: principal component analysis

PCR: polymerase chain reaction

PDA: potato dextrose agar

PDB: potato dextrose broth

PIS: precursor ion scanning

ppm: parts-per-million

psi: pounds per square inch

PTFE: polytetrafluoroethylene

QC: quality control

QqQ: triple quadrupole

Q-TOF: quadrupole time-of-flight

RAL: resorcyclic acid lactone

RF: radio frequency

RP: reverse-phase

RRHD: rapid resolution high definition

RT: retention time

sp.: species (singular)

spp.: species (plural)

TAFC: N, N', N''-triacetylfusarinine C

TEF1α: translation elongation factor 1- α

TIC: total ion current

UHPLC: Ultra-high-performance liquid chromatography

UPLC: ultra-performance liquid chromatography

List of Appendices

Chapter 1: Introduction

Fungi are a diverse group of microorganisms that grow in virtually every habitat on earth. It is estimated that up to 9.9 million fungal species exist, although fewer than 100,000 have been identified.[1, 2] These species have a number of roles in relation to humans. They may be edible mushrooms, yeasts that assist in food fermentation, pathogens that infect crops or livestock, or infectious agents in human diseases.

Ascomycetes are the most abundant class of fungi, comprising over 64,000 unique species.[3] These are considered the "higher fungi", due to their complex structures and elaborate reproductive cycles. Many are described as "filamentous" because they grow by extending segmented filaments called hyphae. This class shows extreme structural and functional diversity, encompassing both pathogens and beneficial fungi. Ascomycetes are also a rich source of natural products. Natural products are compounds produced by living organisms which have relevance to many fields, such as medicine, agriculture, and commerce. For example, penicillin derived from the fungal species *Penicillium chrysogenum* (formerly *Penicillium notatum*) is one of humanity's most important natural product discoveries due to its potent antimicrobial activity.[4, 5] Taxol (which has the product name Paclitaxel) is an anti-cancer drug sourced from Pacific yew tree bark that has been crucial in chemotherapy for breast and ovarian cancers.[6] Certain natural products are also commercial commodities, such as cannabinoid-containing oils and extracts, which are a growing sector of the Canadian cannabis industry.[7] Many natural products are used as-is; they are directly extracted and purified from their natural source. Others are used as the base for semi-synthetic products or have fully synthetic versions.[1] Penicillin is extracted directly from the liquid media of *Penicillium* cultures, while taxol is now fully synthesized.[1]

Microorganisms contribute significantly to the list of important natural products. Over 75% of known antibiotics are produced by microorganisms, including 20% produced by the filamentous fungi.[1] Although a wealth of important natural products do arise from microbes, many microorganisms remain unexamined. It is estimated that less than 1% of microbes that exist in the world have been successfully cultured in the lab, which leaves many undiscovered natural products with potentially novel structures and activities.[1] It is important to screen the natural world for its products, not only because of the potential aid it may provide, but also to gain knowledge about their function within an organism or environment.

While there are many benefits to studying natural products, the process presents challenges. Crude extracts from organisms are complex mixtures with many constituents that are often unknown. There may also be hundreds of unique compounds within a sample that are worthy of investigation. Therefore, choices made concerning fungal culturing, sample collection, analytical techniques, and data analysis are crucial for examining natural products.

1.1 Culturing Fungi

Some key issues arise in studying fungi in a laboratory — namely, consistency in maintaining fungal cultures, and ensuring those cultures are free from other organisms. Maintenance and subsequent analysis of fungi in a laboratory setting can be done in a variety of ways, the easiest and most inexpensive being culturing in liquid media.[8] Fungi are commonly inoculated in liquid media in Roux bottles — flat bottles that maximize surface area, and therefore nutrient uptake — and are left to ferment for a specific amount of time (see Figure 1, below). Fungi form a mat of mycelia across the surface of the liquid and produce natural products as they grow.[8] Many natural products are exuded into the nutrient broth, however, there are some that remain exclusively within the cells. Unfortunately, still liquid fermentation can suffer from inhomogeneity

to some degree as the culture grows. Nutrients are used up more rapidly nearest to the culture, forming a nutrient gradient in the broth. Nutrients at the media-culture interface are replaced by diffusion from other more nutrient-rich parts of the broth, but this occurs slowly. Cultures can therefore be optionally shaken or aerated throughout fermentation to allow for more complete nutrient uptake, because these methods reduce the nutrient gradient that forms in still culture.[8] While fermentation in liquid media is simple and inexpensive, this method is not especially desirable for rapidly screening many cultures simultaneously, because liquid media is more time-consuming to inoculate and extract, and cultures occupy a large volume of space.

Figure 1. Fungal culture grown in liquid media. A one-litre Roux bottle filled with 200 mL of Potato Dextrose Broth and inoculated with a fungal culture. The bottle is sealed with a sterile foam plug to allow gas exchange to occur. The bottle's shape provides a larger growth area than a beaker or flask of similar volume, and therefore provides a better environment for the production of natural products.

An alternate method for growing fungal cultures is to ferment on solid agar media in Petri plates. These use a small volume of media and take up less space than Roux bottles. The growth of fungi across the plates is inhomogeneous similarly to still liquid media, however, the issue can be circumvented by sampling from several different points around the colonies (see Figure 2,

below).[9] Most filamentous fungi grow radially on solid media, meaning that from the point they are inoculated on a plate, they grow outwards in a ring shape until they reach the outer edge of the plate or another colony of fungi. Therefore, the centre of the colony is always in a different stage of growth compared with the outer edge and may be exuding more or less of certain secondary metabolites into its environment. Sampling from several different spots across the plate ensures that a more representative mixture of growth stages and cell exudates is analyzed.

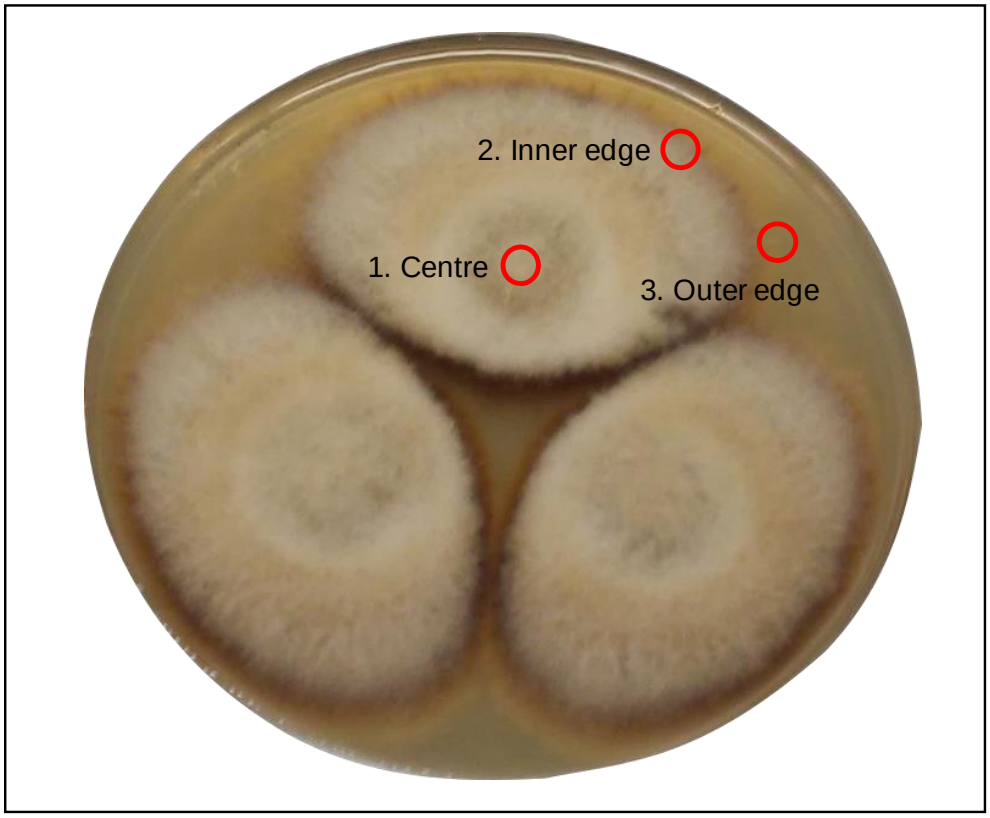


Figure 2. Example of three-point inoculation. A fungal pathogen that was inoculated on Potato Dextrose Agar and grown for 21 days. Sampling regions for rapid screening of natural products are highlighted in red.

To keep cultures sterile, media must be autoclaved before fungal cultures can be inoculated. Autoclaving involves heating materials at a high temperature within a chamber, usually 121 °C, and putting the contents of the chamber under vacuum pressure. The combination of heat and

pressure kills any microorganisms present. Care must then be taken after autoclaving to inoculate only the organism of interest, usually by working in a sterile environment such as a biosafety cabinet.

Metabolites are categorized as either primary or secondary. Primary metabolites are those that are directly involved in processes to maintain life such as catabolic by-products like lactic acid or ethanol, or molecules used in anabolism like amino acids and sugars. Secondary metabolites are compounds that do not serve a direct function in maintaining life.[10] They instead are often made as defense mechanisms or to assist in invading a host. Most natural products in use fall into this latter category. Many secondary metabolites arise from intermediate products of the Krebs cycle. For example, acetyl coenzyme A molecules are linked together to form mevalonic acid, which is an important intermediate in terpene biosynthesis.[8] Terpenes are important plant secondary metabolites that play a role in defense strategies.[11] Other secondary metabolites are formed from combinations of primary metabolites that do not enter the Krebs cycle, such as cyclic peptides formed from amino acids.[8]

Different species prefer different combinations of nutrients and may produce different ratios of products depending on their inputs. An example of altering natural product yields by manipulating nutritional inputs can be seen when growing *Fusarium oxysporum*. The production of enniatin mycotoxins by this species increases drastically upon supplementation of 20 mM L-valine.[12] In the lab, specific media is chosen depending on need. For fungal culture isolation, potato dextrose agar (PDA) is the most widely used. It is nutritionally dense and promotes sporulation as well as the formation of other morphological structures that assist in identification.[13] It is also suitable for growing many species of fungi, making it a good choice for unidentified species. While PDA is broadly useful, some species prefer drier environments. Xerophilic fungi like *Penicillium*

and *Aspergillus* grow better on Czapek yeast autolysate because of its low water activity.[14] Additionally, media usually must also be supplemented with trace amounts of metals, like copper and zinc.[15] Historically, media were made with tap or well water, which contain trace elements that help fungi grow more uniformly. The widespread use of distilled water in media preparation has necessitated supplementation with trace metals.[15]

Another consideration in screening fungi for natural products is at which point in the growth cycle to take samples. The production of secondary metabolites begins after an essential nutrient has been depleted from the medium – usually either nitrogen or phosphorus, depending on the species.[8] The concentration of secondary metabolites reaches its peak when the dry weight of the fungal culture has plateaued, although it may not remain at this maximum for long. Unused secondary metabolites can be recycled for their nutritive components, or they may naturally degrade if left for too long. There is no prescribed time at which to sample cultures for their natural products because growth rates vary so widely across fungal species. Cultures contain secondary metabolites after their exponential growth phase, therefore this time point must be assessed for individual species of interest.

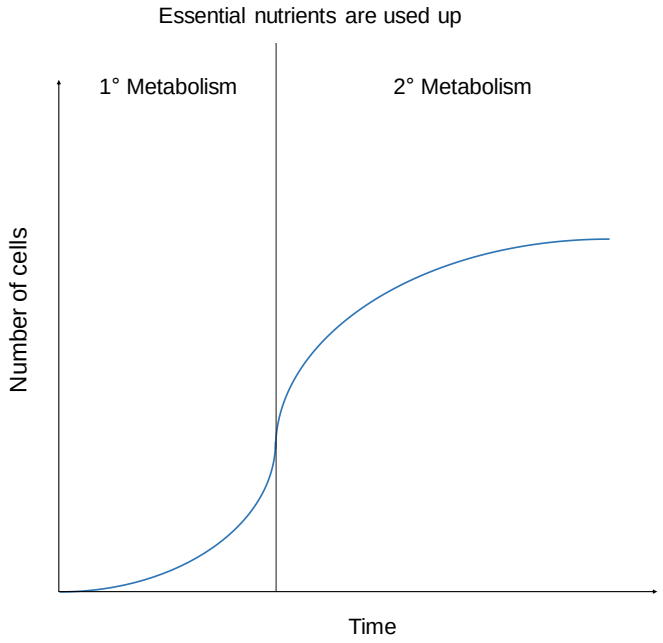


Figure 3. Growth curve of a typical fungal culture. Cells grow exponentially in the primary metabolism phase, then slow down in the secondary metabolism phase, after essential nutrients are depleted.

1.2 Analytical Techniques

1.2.1 Ultra-High-Performance Liquid Chromatography (UHPLC)

Liquid chromatography (LC) is the process of separating components of a solution based on their individual polarities. In its simplest form, it consists of a mobile phase composed of solvents and a stationary phase usually composed of silica that has been functionalized with groups of varying polarities.[16] The most common mode of chromatography in the field of natural products research is reverse-phase (RP) chromatography, which employs a polar mobile phase and a non-polar stationary phase made of small silica beads functionalized with hydrocarbon chains.[16] A complex mixture of analytes is injected into the mobile phase, which flows along the stationary phase. Analytes that are more alike in polarity to the stationary phase interact more strongly with it and are retained for longer, therefore polar compounds elute more rapidly while less polar

compounds elute later. The mobile phase is often a blend of polar solvents, some of the most common being water, acetonitrile, and methanol.

Compounds may be separated by either isocratic or gradient elution. In isocratic separation, the ratio of solvents is held constant over the course of the run. This is useful for separating compounds that interact weakly with the stationary phase or for mixtures of fewer than ten compounds.[17] In gradient elution for reverse-phase chromatography, the concentration of the less polar solvent increases over time, which elutes less polar analytes interacting with the stationary phase. Gradient elution has higher peak capacity and allows for rapid analysis of highly complex mixtures.[17]

High-performance liquid chromatography (HPLC) employs a stationary phase that is composed of small silica beads functionalized on the exterior surface. These beads are tightly packed into a metal column and the mobile phase is pumped through the column. Mixtures of compounds that are injected at the start of the column are separated by polarity by the stationary and mobile phases, then elute individually at the end of the column. Analytes are identified at the end of the column by a detector, oftentimes a UV or mass detector.

Ultra-High-Performance Liquid Chromatography (UHPLC) is a high-resolution separation technique that was initially developed and trademarked by Waters Corporation as "Ultra-Performance Liquid Chromatography" (UPLC). It improves analyte resolution compared to standard HPLC by applying higher pressures of up to 15,000 psi and by employing a stationary phase with smaller diameter packing materials.[18] UHPLC is useful for analyzing the complex mixtures that are often examined in metabolomics. Separation by LC prior to ionization and mass detection also reduces matrix effects and allows for the separation of isomers that would be considered identical in a direct-infusion experiment.

1.2.2 Mass Spectrometry

Mass spectrometry allows us to measure the mass of compounds. At the single-compound scale, mass is denoted in Daltons (Da), also known as standardized atomic mass units (amu), which are equal to one twelfth the mass of a carbon atom (1 amu = 1.66 x 10^{-27} kg).[19] However, the true unit of measurement of mass spectrometry is the mass-to-charge ratio (*m/z*). No matter the type of mass spectrometer, analytes must be ionized before they can be detected, whether by addition or subtraction of charged groups such as electrons, protons, or other charged adducts. Notation for masses is given by $[\text{Mass} \pm \text{group}]_{\text{charge}}$. Commonly, protonated, sodiated, and ammoniated ions are visible within a spectrum, denoted by $[M+H]^+$, $[M+Na]^+$, and $[M+NH_4]^+$, respectively. There may also be different ion forms of one analyte that occur concurrently within a spectrum. Analyte detectors rely on the charge of gas-phase ionic species to trigger a charge differential at the detector that produces a signal. Mass spectrometers are often coupled with HPLC systems, because they separate compounds and are easily coupled with electrospray ionization. An Orbitrap Q-Exactive mass spectrometer with an electrospray ionization source was used for all analyses herein, which uses alternating current (AC) to set *m/z* windows to include ions for analysis in the mass detector. Components of the Orbitrap Q-Exactive mass spectrometer are described below.

1.2.2.1 Heated Electrospray Ionization (HESI)

Electrospray ionization (ESI) is an atmospheric pressure ionization technique first developed in 1984 by Fenn and Yamashita and has become one of the most common ionization methods, especially for studying biological molecules.[20, 21] It is ideal for many analytes, from small molecules to large proteins. It is a "soft" ionization technique, meaning compounds are not fragmented by the ionization process, or fragmentation is minimal. This is in contrast with some of the older ionization techniques, such as electron impact ionization, which fragments the analyte

in the process of giving it a charge.[19] The "softness" of ESI allows the determination of *m/z* values for intact ions, which can optionally be fragmented later in the instrument. This is especially desirable for applications where the contents of a mixture are unknown, because chemical formulas can be determined from high-resolution *m/z* values.[22]

ESI can operate in both positive and negative ionization modes. Analytes may ionize preferentially in one mode over the other, depending on their structures. Analytes with acidic groups are suitable for negative ionization mode, because their acidic protons are easily removed by hydroxyl groups present in the basic capillary environment. Basic groups such as amines that can accept a proton are good candidates for positive mode, where there are excess protons acidifying the liquid in the capillary. A benefit of ESI is that it tends to produce multiply charged ions — especially for larger biomolecules like proteins — which gives them a *m/z* within the operable range of instruments.[19]

Analytes in solution are pumped through a small metal capillary which has an electrical current running through it. In positive mode, a power current draws electrons away from the solution in the capillary and forces the electrolysis of water. By-products of water oxidation are seen in equation **1**, below. Oxygen gas and protons are formed as water is electrolysed. Free protons can attack basic groups and form hydrogen adducts with analytes, simultaneously giving them a positive charge. In negative mode, electrons are guided towards the capillary. Excess electrons reduce water to hydrogen gas and free hydroxyl groups that remove hydrogens at analytes' acidic sites, giving those sites negative charge. By-products of water reduction are presented in equation **2**. Analytes may also form adducts with sodium and potassium ions that are shed from glassware during sample preparation, or with ammonium ions that may be present as solvent additives.[23]

$$2H_2O \rightarrow 4H^+ + O_2 + 4e^- \quad \textbf{(1)}$$

$$2H_2O + 2e^- \rightarrow H_2 + 2OH^- \quad \textbf{(2)}$$

In both ionization modes, electrochemical distortion at the mouth of the capillary causes the formation of a Taylor cone. Droplets form at the tip of the Taylor cone as a nebuliser gas (commonly nitrogen gas) is blown in parallel to the capillary.[24, 25] This can be seen below, in Figure 4. Once ion-containing droplets are formed, they must be desolvated before entering the detector.

There are two main theories explaining how small molecules become desolvated from droplets, including the Ion Evaporation Model (IEM), and the Charge Residue Model (CRM)[26]. The IEM, mostly pertaining to analytes under 1000 Da, posits that as droplets evaporate, charge accumulates on the outside of the droplet such that it becomes more energetically favourable to expel the desolvated ion from the central droplet than to retain it.[27] For large analytes such as large globular proteins, the CRM applies.[28] Droplets undergo jet fission, becoming smaller and more numerous due to evaporation and electrostatic repulsion. Eventually each droplet contains only one analyte molecule and all liquid evaporates, depositing charges or adducts previously contained in the droplet on the surface of the analyte.[26]

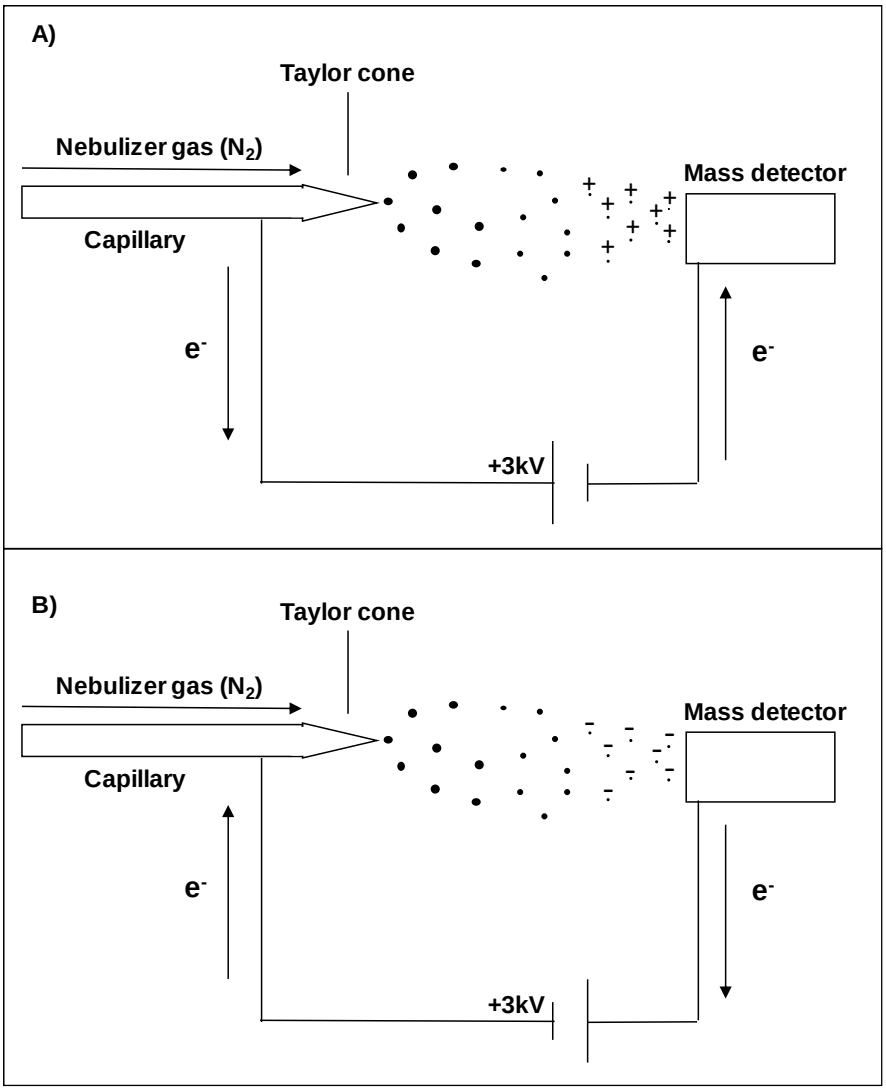


Figure 4. The formation of ions by electrospray ionization. A) Positive mode of electrospray ionization. A current withdraws electrons from the capillary, allowing for oxidation of water in the capillary. B) Negative mode of electrospray ionization. A current brings excess electrons to the capillary, allowing reduction of water. In both modes a nebuliser gas helps the stream of solvent exiting the capillary to form droplets. Droplets evaporate as they travel toward the mass detector, exposing ions within that can be detected

ESI is commonly coupled with a liquid chromatography system, which allows for separation of compounds based on polarity before they are sent to the ionization source and then to the mass detector. Standard reverse-phase solvents are used, but it is common to include

additives that improve ionization.[29] Popular volatile additives include formic acid, trifluoroacetic acid, ammonium formate, ammonium acetate, and acetic acid.[29]

1.2.2.2 S-Lens, Flatapoles, Quadrupole Mass Filter, and C-trap

Once an analyte has been ionized, it must be brought to the mass detector. Atmospheric gases present a large obstacle for ions to bypass, therefore, mass detectors are kept under high vacuum to reduce that obstacle. In an Orbitrap mass detector coupled with an ESI source, ions are formed at atmospheric pressure, then must travel to the mass detector which is kept at high vacuum.[24] To reduce the workload on a single pump, atmospheric pressure ionization (API) techniques use a differentially pumped vacuum system. Roughing pumps maintain two areas of decreasing pressure, ranging from 2 mbar near the ionization source to 3×10^{-5} mbar in the quadrupole region.[24] This makes it easier for subsequent turbo pumps to maintain a vacuum of 1×10^{-9} mbar within the mass analyzer.[24]

Ions formed at atmospheric pressure are drawn into a small metal capillary at the interface of atmospheric pressure and vacuum pressure. The capillary is heated to improve solvent evaporation and a collision voltage is applied to accelerate ions into one another, further desolvating droplets by gentle collision.[30] This sampling interface replaces the skimmers and cones often used in other instuments.[31] Ions are then guided from the capillary toward the S-lens (see Figure 5, below).

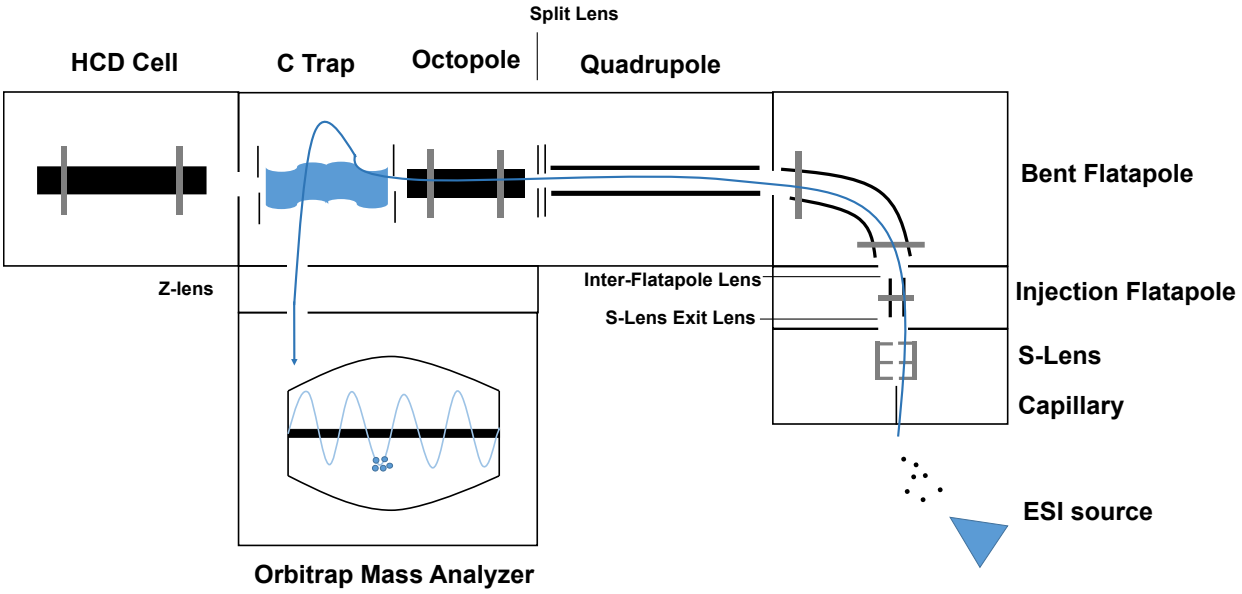


Figure 5. Schematic of ion movement through an Orbitrap Q Exactive mass spectrometer. The path of ions through the instrument is highlighted in blue.[32]

The S-lens is composed of stacked metal rings that focus ions into a beam using radiofrequencies (RF).[33, 34] By applying RF voltages of opposing phases to alternating rings, ions are pushed forward through the apertures of the rings.[34] The ion beam travels towards the injection flatapole, which assists with further desolvation of larger analytes, then through to the bent flatapole – a specialized multipole component that operates similarly to a quadrupole.[32, 35] The curve of the bent flatapole allows ions to be guided and focused again as they round a corner, while any stray neutral contaminants travel straight forward and collide with the wall of the instrument.[24] Several additional lenses are placed along the ion path, to assist ion transmission. Ions move through a hyperbolic quadrupole with a gated end, then through a short octopole. Ions eventually reach the C-trap, which is an RF-only modified linear ion trap with a slight curve in its shape.[36] Ions can travel directly through, or may be stored in the C-trap for a short duration to allow downstream analysis in the Orbitrap to occur uninterrupted. From the C-trap, ions are injected into the Orbitrap mass analyzer, where their *m/z* is detected.

1.2.2.3 Orbitrap Mass Spectrometer

The Orbitrap mass detector is a newer ion trap-type mass analyzer that operates by Fourier transform, which was developed by Thermo Scientific in 2005.[37] The Orbitrap is a high-resolution mass detector capable of resolution up to 140,000.[32] Resolution is the ability of an instrument to distinguish *m/z* values, or a measurement of how narrow a peak is. It is given by equation **3** below, where Δ*m/z* is the full width of a peak at half of its maximum (FWHM, shown in Figure 6).

$$R = \frac{m/z}{\Delta m/z} \quad \textbf{(3)}$$

Resolution is not constant across the mass scanning range. Resolving power is highest for lower *m/z* values and it decreases as the *m/z* of a compound increases.[32]

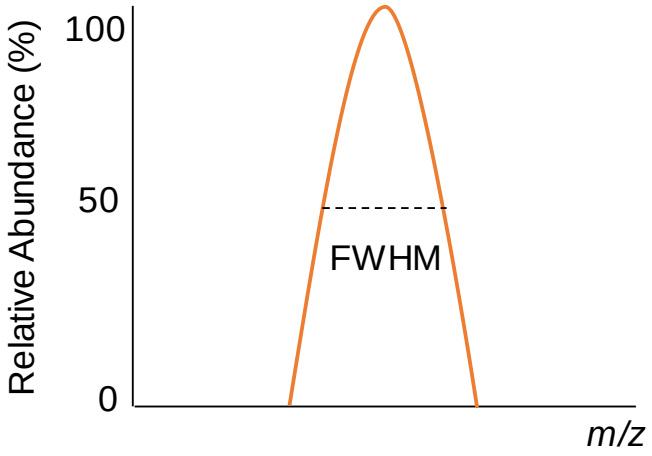


Figure 6. Mass resolution using full width at half maximum (FWHM).

Another important metric in mass spectrometry is mass accuracy, which is a measure of how close the experimentally measured mass is to the theoretical mass of a compound. Mass accuracy is measured in parts-per-million (ppm) and is described by equation **4**, below. Milli-mass units (mmu) may sometimes be used to describe error for a single measurement.[38]

$$mass\ accuracy\ (ppm) = \frac{Measured\ mass - \ Exact\ mass}{Exact\ mass}\ x\ 10^6 \qquad \textbf{(4)}$$

Orbitrap mass spectrometers routinely achieve excellent mass accuracies, usually within 0.5-1 ppm.[39] For comparison, quadrupole time-of-flight (Q-ToF) and triple quadrupole (QqQ) instruments can achieve 3-5 ppm accuracy and linear ion traps (LITs) can get to 50-200 ppm.[39]

The Orbitrap mass detector is constructed of a negatively charged spindle with a positively charged shell encasing the outside.[40] The shell has a small gap in the centre where a current differential is applied to both sides. Ions are injected into the detector perpendicularly to the spindle which gives them ample excitation to orbit the spindle's axis.[41] The trajectory of an ion about the axis is governed by its *m/z*, therefore specific *m/z* values group together in packets. These packets orbit around the spindle while also moving back and forth across it (see Figure 7, below).

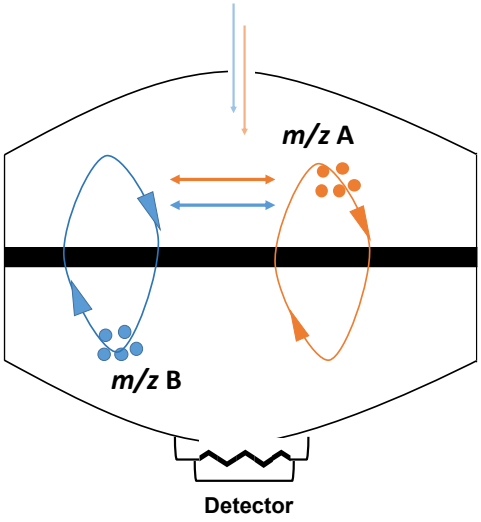


Figure 7. Analyte movement through an Orbitrap mass analyzer. Ions are injected from the C-trap perpendicularly to the spindle via the gap in the outer electrode. Here, two different *m/z* values orbit about the central spindle of the Orbitrap mass detector. Ions also pass back and forth across the gap to form an image current across the resistor, and the frequency of oscillation corresponds to a specific *m/z*.

Ion packets are detected as they move back and forth across the differentially charged gap at their specific frequency. Analyte oscillations are recorded as the decay of a signal over time. The time domain is Fourier transformed (FT) into a frequency domain, which is then translated into a unique *m/z*. Peak resolution is improved by allowing longer scanning times for transients.[40]

An interesting aspect of data acquired with FT-type mass spectrometers is that they do not have traditional background noise, as with other instruments. This is due to the difference in ion detection it has from other ion detection systems. Many instruments produce a signal when ions hit a detector, therefore they are prone to some level of neutral or stray ionic compounds triggering a signal when they hit the detector, which causes a constant background signal or "noise". Because the only trigger for FT-type mass spectrometers is the charge of *m/z* packets, they are not subject to this noise, although there may still be ubiquitously expressed compounds from the mobile phase, especially for LC-MS.[42]

LC-MS data can be visualized in a few different ways, usually in some variation of *m/z* as a function of time. Total ion current chromatograms (TICs) show the summed intensity of all ions detected at each time point. Base peak chromatograms show only the most intense ion at each time point. Extracted ion chromatograms (EICs) show the intensity of a specific *m/z* at each time point. Examples of these are all shown below, in Figure 8.

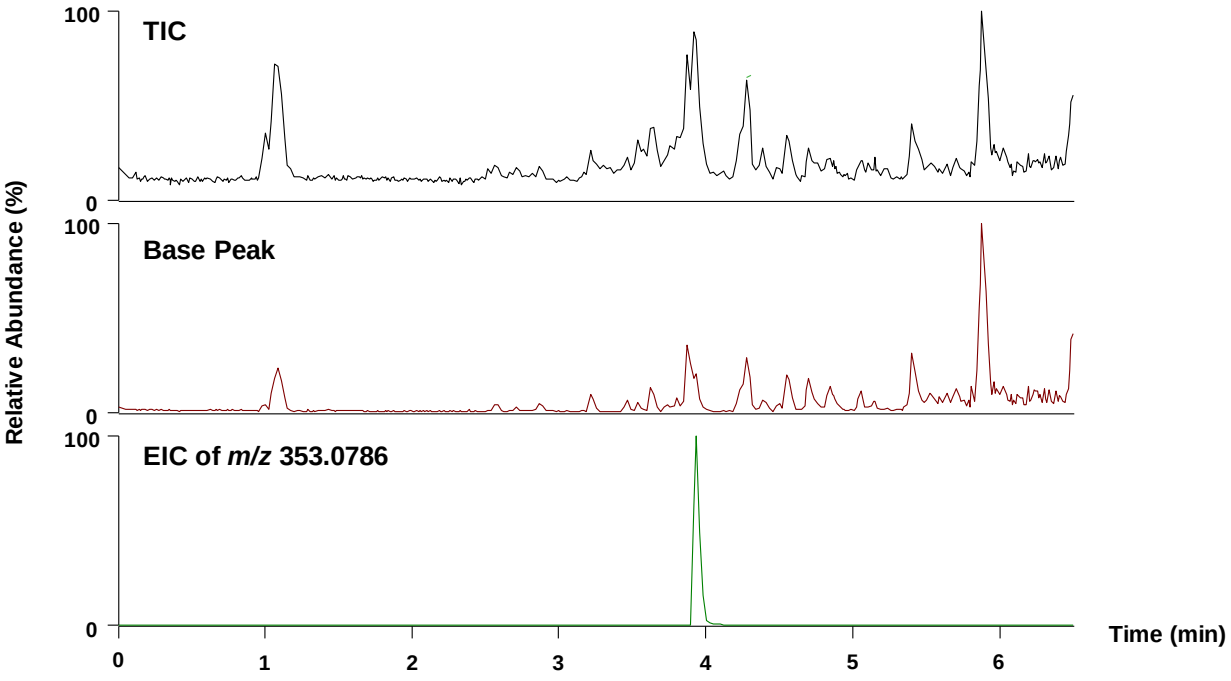


Figure 8. Different chromatogram representations of one LC-MS raw file. Positive ionization mode LC-HRMS file of methanol extract of fungal endophyte E-184. Top: total ion current chromatogram. Middle: Base peak representation. Bottom: extracted ion chromatogram of *m/z* 353.0786.

1.2.2.4 Higher Energy Collisional Dissociation Cell

In the Orbitrap, ions can be analyzed intact or can be sent to be fragmented by higher-energy collisional dissociation (HCD) to gather more information on their structures. HCD is termed such because the HCD cell is held at a higher RF voltage than standard collision-induced dissociation (CID) cells (2500 V compared to 1500 V in a regular CID setup).[43] In HCD, compounds are fragmented by applying a "collision voltage" to the analyte ions, causing their acceleration into a neutral gas that is unaffected by the applied voltage, usually nitrogen.[43] Collisions cause analytes to heat up and gain energy that becomes distributed among their bonds.[44] Once a molecule has absorbed enough energy to overcome the activation energy of the lowest resistance bond, that bond will dissociate, forming at least one ionic fragment and one neutral

fragment.[44] Neutral fragments (aka. neutral losses) are not seen by the mass detector but can be inferred from mass differences between fragment ions.

One issue in fragmentation is that larger molecules have more bonds than smaller ones; therefore, they absorb more energy than smaller molecules do before they can overcome activation energies and dissociate into fragments. If the same collision energy is applied to a small molecule and a large one, the small one may fragment into several pieces, while the larger one may not fragment at all. To improve fragmentation across a wide mass range, the HCD cell employs normalized collision energy, where the amount of energy a molecule receives is relative to its *m/z*.[43] The collision voltage is set as a percent of 5V and is adjusted for different *m/z* values so analyte ions experience the same voltage, regardless of *m/z*.[45] Methods can make use of one specific collision energy or can optionally fragment the same set of ions at multiple distinct collision energies, termed stepped collision energy.[46] Precursor ions are fragmented at multiple energies, sequentially, then all fragments are sent to the mass detector for analysis.[47] This technique is desirable for fragmenting unknown metabolites, as it improves the number of fragments that can be searched in databases.[46]

In Q-TOFs and triple quad instruments, the mass analyzer can double as a collision cell, however this is not true for Orbitraps.[41] The HCD cell is a workaround for this issue. A small slice of ions is measured in the Orbitrap to determine the most intense ions. These most intense ions are also still present in the C-trap and they are then sent to the HCD cell to be fragmented.[41] Fragments are sent back into the Orbitrap to be analyzed. While ions are fragmented, incoming ions from the ESI source are held in the C-trap so fragments can be correctly assigned to their precursor ions.

MS/MS methods for untargeted metabolomics analysis are often data-dependent acquisition (DDA) methods, where the most intense ions in a scan are chosen for fragmentation.

Specific parameters can be adjusted for DDA experiments. The use of automatic gain control (AGC) was first developed for use with FT-ICR instruments, but it has since been assimilated into Orbitrap models as well.[36, 48, 49] This setting controls the number of ions that accumulate within the LIT before an MS or MS/MS event is triggered. This is closely tied to the maximum injection time (max IT), which is the maximum amount of time permitted to elapse before ions are sent to the mass analyzer. MS/MS events will occur when one of these parameters is fulfilled, regardless of the status of the other.[50]

1.3 Metabolomics

1.3.1 High-resolution LC-MS in metabolomics

Metabolomics is the study of the entire profile of small molecules (i.e. molecules weighing less than 2000 Da) produced by an organism, also known as the metabolome.[51] Along with genomics and proteomics, it is a useful way of examining biological systems.[52] It has become a popular tool in recent years because many insights can be gleaned from examining both the types of compounds produced and the amounts of those compounds. For example, by determining the relative levels of primary metabolites within a sample one can assess if specific biosynthetic pathways are disrupted by experimental treatments or disease.[53] Screening organisms for secondary metabolites is also prominent in the field of drug discovery, and in learning about how organisms interact with their environments.[1] Screening may occur through either targeted or untargeted methods. In targeted methods, a predetermined list of compounds is searched for within the samples. These methods are often paired with standards to allow quantification. Untargeted methods usually acquire data based on ion intensity, which gathers information on all compounds within a sample that meet certain intensity parameters.

Analysis by high-resolution mass spectrometry (HRMS) is one of the primary techniques for metabolomics studies, as it allows for acquisition of data with high mass accuracy, i.e. data that are very close to the theoretical mass of a compound.[54] Highly accurate data are necessary for resolving chemical formulas of unknown analytes.[22]

All elements have a different mass defect, meaning their mass is slightly lower than the sum of the mass of their protons and neutrons. This occurs because when protons and neutrons come together in an atom's nucleus, the process is exothermic. Energy is emitted and is ultimately lost as mass, which causes unique mass defects for each element, and for each elemental isotope that may arise. A compound will therefore have a mass defect that is indicative of the number and type of atoms that make up its composition. Chemical formulas of unknown compounds can then be determined from the accurate mass of an analyte, if the resolution is high enough. A low resolution instrument that only distinguishes *m/z* values to one decimal place would not allow us to see a difference between the isobaric compounds glutamine (*m/z* 128.12922) and lysine (*m/z* 128.17228), whereas this difference would be apparent on a high-resolution instrument.

The mass sum of all atoms in a compound is fairly unique, especially for smaller compounds.[55] Mass spectral analysis software often has functions that can calculate chemical formulas from high-resolution *m/z* values. These provide a list of potential chemical formulas based on its internal calculations, however, manual assessment of the formulas is still necessary to ensure the mass accuracy for a predicted formula is acceptable, and that the formula is viable.

HRMS also allows us to see the fine isotope structures of analytes. Many compounds have a natural occurrence of stable isotopic versions of themselves. Carbon-13 exists ubiquitously at 1% abundance, therefore a carbon-containing compound has a 1% chance for each individual carbon to be occupied by a ^{13}C atom. These isotopic peaks can be observed at constant mass

differences from the monoisotopic mass. Other elements such as chlorine, sulfur, and selenium also present characteristic isotope patterns that can indicate which elements are part of a particular compound. For example, chlorine has a distinct isotope pattern where the monoisotopic peak of an analyte is approximately three times higher than the *M + 2* peak and they differ by 1.997 Da, (Figure 9, below).

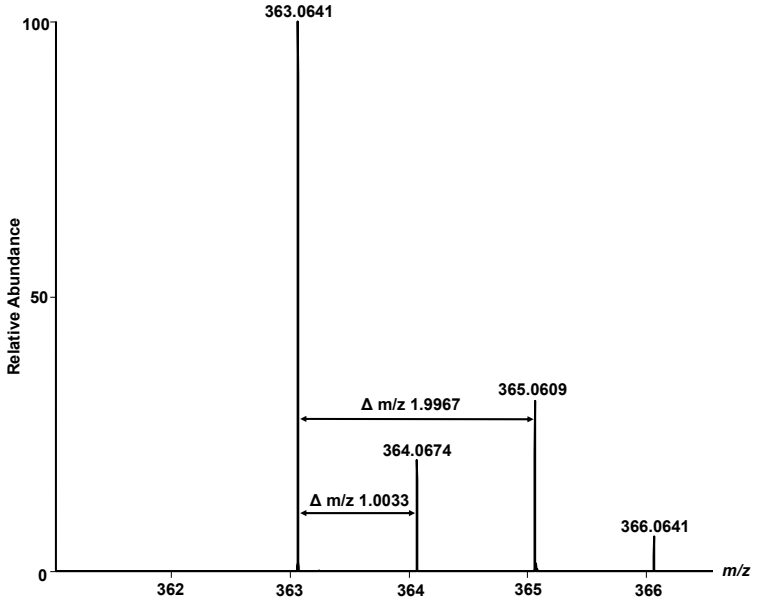


Figure 9. Fine isotope structure of a chlorine-containing compound. The compound has a *m/z* of 363.0641, and the spectrum was acquired in negative ionization mode. The first ^{13}C isotope can be seen at a *m/z* 364.0674, while the ^{35}Cl isotope is at *m/z* 365.0609 at approximately 30% relative abundance.

When ionizing by electrospray ionization, often multiply charged ions are formed, therefore there may be ions present that have different masses but share the same *m/z*. Charge states are indicated by the spacing between isotopic peaks. An ion with a higher charge state will have smaller mass differences between its isotopic peaks than an ion with a lower charge state, therefore they are discernable from one another.

1.3.2 Data processing in metabolomics

One of the main roadblocks in the study of metabolomics and natural products discovery is determining previously identified compounds within a sample in the beginning stages of analysis, a process that is termed dereplication.[56] Although thousands of natural products have been isolated and identified, databases containing these compounds are often incomplete or do not provide adequate information to identify unknown compounds. Metabolomics datasets are often massive because thousands of features can be detected in a single sample. Consequently, manual examination of metabolite lists is not an option for many studies. Therefore, many strategies have emerged to deal with large metabolomics datasets.

1.3.2.1 xcms

There are several packages developed for the software program R that assist in handling metabolomics data. R is a popular tool for data processing, as it is free to use, powerful, and has a large community for support.[57] The first step in analyzing HRMS data is to determine all *m/z* values that have been detected in raw files, and to determine which samples produced those peaks. The package *xcms* "reads" HRMS files and identifies all features that fulfill certain user-adjusted parameters.

Before raw files are read into R, they must be converted from profile mode into centroid mode. Most vendor data acquisition programs collect LC-MS data in profile mode, meaning data are collected in a continuous manner.[58] These files contain many data points and are advantageous because they show the entire shape of a peak. However, these data are often too dense for further processing steps. Peaks in centroid mode are indicated by a single data point that denotes *m/z* and

intensity, which reduces file sizes drastically and in turn speeds up processing times (Figure 10, below).

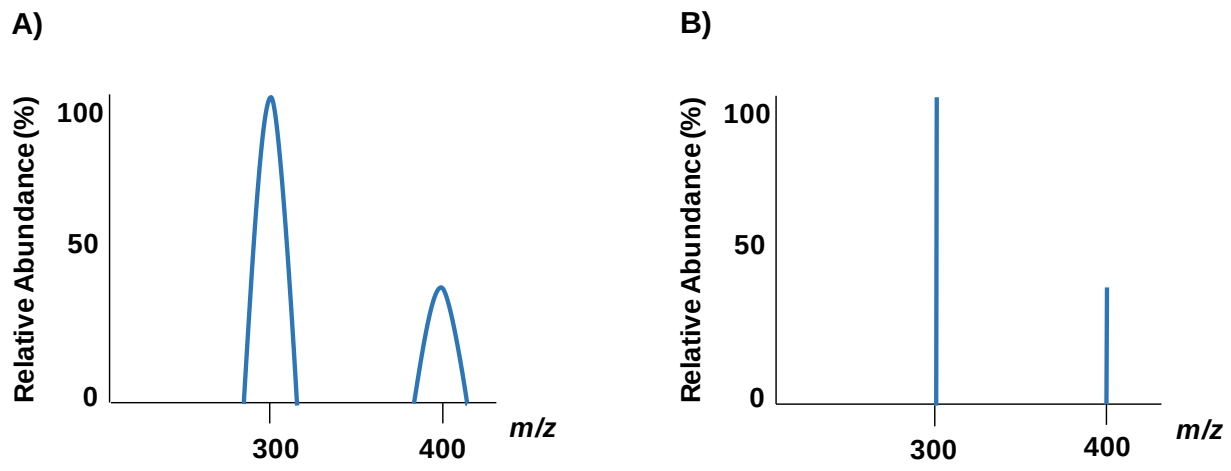


Figure 10. Comparison of profile and centroid mode mass spectra representations. A) Profile mode. B) Centroid mode.

The *xcms* package has several functions that help convert raw files into peak lists. Briefly, the centroided data files are analyzed and all features that are detected within the files are compiled into a peaklist. There are several filtering parameters to include or exclude peaks, which are described in detail in Appendix A. Afterwards, peaks are grouped, so features that are present in multiple samples are identified as being one feature rather than many. Then, another function looks through the raw files once again to find features in the peaklist that are present, but below the threshold provided for initial peaklist compilation. The peaklist output from these functions may then be subjected to statistical tests, which are detailed below.

1.3.2.2 Statistical testing

In metabolomics analysis, one often wants to determine similarities and differences among samples. There are several statistical tests one can perform to identify which variables are

statistically different. Variables in the case of metabolomics datasets are the measured levels of features that have been detected.

Principal component analysis (PCA) was first developed in 1901 by Karl Pearson as a means of extracting the most pertinent variability within multivariate datasets.[59] It has become useful for dealing with the large, multivariate datasets acquired in metabolomics, because it condenses data variability into fewer dimensions, making it easier to assess patterns.[60] It is also unsupervised, so any patterns that are discerned are not due to imputed bias, such as information about treatment groups.[60] It is a statistical test that transforms sets of potentially correlated variables into sets of uncorrelated variables called principal components.[60] When samples are plotted along the principal components, samples that are closer together are more similar, while samples that are farther apart are more different.

Concurrently alongside PCA calculations, factor loadings are generated. These are the contributions of each factor to the distribution of samples within the principal components. In metabolomics, factors are individual features identified by LC-MS. Factor loadings are plotted separately from samples, but are plotted within the same principal components as samples. They are represented as arrows that point towards the sample or group that makes that feature in the highest abundance. If patterns or groups arise visually in the PCA plots due to differences among variables, one may wish to determine if those differences are statistically significant. There are several statistical tests available to determine if the means of two groups are significantly different, but for metabolomics, often a Kruskal-Wallis test for two or more groups of non-parametric data is used.[61, 62] This is a test that can be applied to non-normally distributed datasets, which is often a reasonable assumption with metabolomics data. Kruskal-Wallis tests compare the means of the variables of given groups to yield a *p*-value. Usually *p*-values less than 0.05 are considered

significant, although the smaller the *p*-value, the more confident one can be that the means are truly different among groups. One complicating factor, however, is that a certain number of false positives occur when doing multiple comparisons. Benjamini-Hochberg correction is a calculation applied to *p*-values acquired through statistical comparison testing that reduces the false discovery rate.[63] It is performed to ensure false positives do not influence the analysis.

1.3.2.3 Molecular Networking

When applying PCA and other statistical tests, only the first MS level is used, but many additional insights can be gleaned from examining the second MS level. In tandem mass spec experiments, compounds are fragmented into pieces. Structurally related compounds will fragment in similar ways when they undergo MS/MS. There are several methods that use this characteristic of tandem mass spec experiments to gather information about classes of compounds. Triple quadrupole (QqQ) instruments can perform neutral loss scanning (NLS) and precursor ion scanning (PIS) MS/MS experiments.[64] In NLS, the first quadrupole (Q_1) scans for precursor masses, CID occurs in the second quadrupole (q_2), then the third quadrupole (Q_3) scans for neutral losses, for example losses of H_2O or NH_3.[65] In PIS, Q_3 is fixed on a product ion while Q_1 scans for precursor masses that produce that product ion, usually indicative of a functional group.[66] While these methods work for quadrupole-type mass analyzers, true scanning cannot be done on Orbitrap mass analyzers, as NLS and PIS both require two mass analyzers scanning concurrently, offset by a neutral loss or searching for a specific product ion.[67] Therefore, filtering or post-acquisition methods can be used for a similar effect. Diagnostic Fragmentation Filtering (DFF) is a post-acquisition method that examines MS/MS data for particular fragments or neutral losses that are characteristic to a class of compounds.[68] DFF is beneficial as it can rapidly assess data for particular targeted classes and produces a relatively short list of precursor ions. However, the

method is less effective when there are no known or suspected compound classes or fragment ions in a mixture.

Spectral molecular networking circumvents the issue of needing to know product ions or neutral losses by comparing entire tandem mass spectra rather than mining for specific product ions. This technique was developed by researchers from the Scripps Institute and UC San Diego as a way to tackle the issue of compound dereplication in natural product discovery.[69] Often, samples used to study secondary metabolites are complex mixtures containing both known and unknown products. Being able to identify previously discovered compounds early in the analysis saves time and resources and allows for more focus on discovering novel compounds.[56] Spectral molecular networking was later integrated into Global Natural Products Social Molecular Networking (GNPS) - a user-curated database of natural products that allows comparison of tandem mass spectra with publicly shared datasets and annotated spectra.[70]

Briefly, a computer program mathematically compares all spectra it is given and groups them visually based on the similarity of their fragments. All spectra detected within the uploaded data are compared with one another, and with all spectra available in GNPS databases. Users may also optionally compare their data with any publicly available user datasets. The resulting network can be visualized in a program such as Cytoscape. Individual features are represented as nodes, which are connected to other features by lines if their spectral similarity is above a specified value. Features may also be dereplicated and annotated if they share enough fragment ions with spectra in GNPS databases or user files. A detailed explanation of this process is given below, in Section 1.3.3.

1.3.3 Molecular Networking

Analyte fragmentation in the HCD cell yields product ions and neutral losses that are characteristic of those analytes. Fragmentation reactivity is dictated by a compound's structure, therefore, compounds that are structurally similar tend to produce similar product ions.[71] The similarity of compound structures can thus be inferred from the similarity of their fragmentation spectra, if they are fragmented under similar conditions.

Tandem mass spectra may be converted into vectors, where each axis represents the *m/z* of a fragment and values along each axis represent the intensity of that fragment.[72] This can be illustrated with two simple spectra, each containing one fragment (Figure 11, below).

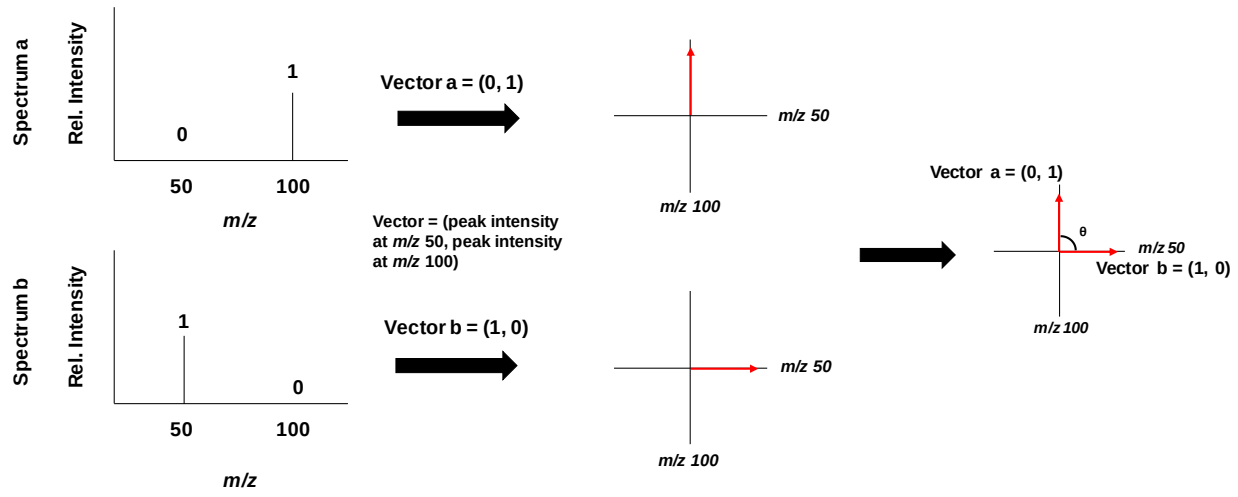


Figure 11. Example of two simple mass spectra being converted into vector representation. The relative intensities of each peak are converted into vector coordinates, which are plotted on a set of coordinate axes. The angle between the vectors, θ, can then be used to compare them.

In this case, there are only two options for fragments, *m/z* 50 and 100. The intensities of those fragments can be translated into vector coordinates, then plotted on axes. Representing spectra as vectors also allows comparison of vectors using a dot product calculation. The result of the calculation, named a cosine similarity score, indicates the degree of relatedness shared by the

vectors.[73] Results closer to 1 indicate a higher degree of spectral similarity and those closer to 0 indicate more difference. Comparison of the spectra from Figure 11 by a dot product calculation is demonstrated below using equation **5**.

$$\cos\theta = \frac{\sum_1^n a_i b_i}{\sqrt{\sum_1^n a_i^2} \times \sqrt{\sum_1^n b_i^2}} \quad (\mathbf{5})$$

Vector values from Figure 11 can be substituted into equation **5**:

$$= \frac{(0 \times 1) + (1 \times 0)}{\sqrt{0^2 + 1^2} \times \sqrt{1^2 + 0^2}}$$

$$= \frac{0}{1}$$

$$= 0$$

A solution of 0 indicates that the two vectors share no similarities, which agrees with the sample spectra. They share no common fragments, therefore they are completely dissimilar. Extending this to experimental mass spectra that have many more fragments with more *m/z* values, vectors become plotted in n-dimensional space. Therefore, vectors are impossible to represent visually, but can still be represented and compared mathematically. Dot product calculations are performed between all vectors to compare all spectra with one another. All resulting cosine similarity scores are used to plot spectra in a network, where each spectrum is represented by a single node. Nodes are connected to one another by lines that indicate that the cosine similarity score used to compare them was above a prescribed cut-off point. Using the spectra depicted above, a network would appear as two unconnected nodes. Imagining two theoretical spectra that are similar enough to have a cosine score cut-off above the set threshold, these would appear as

two connected nodes in a molecular network (Figure 12, below). Any groups of connected nodes within a network are referred to as “clusters”.

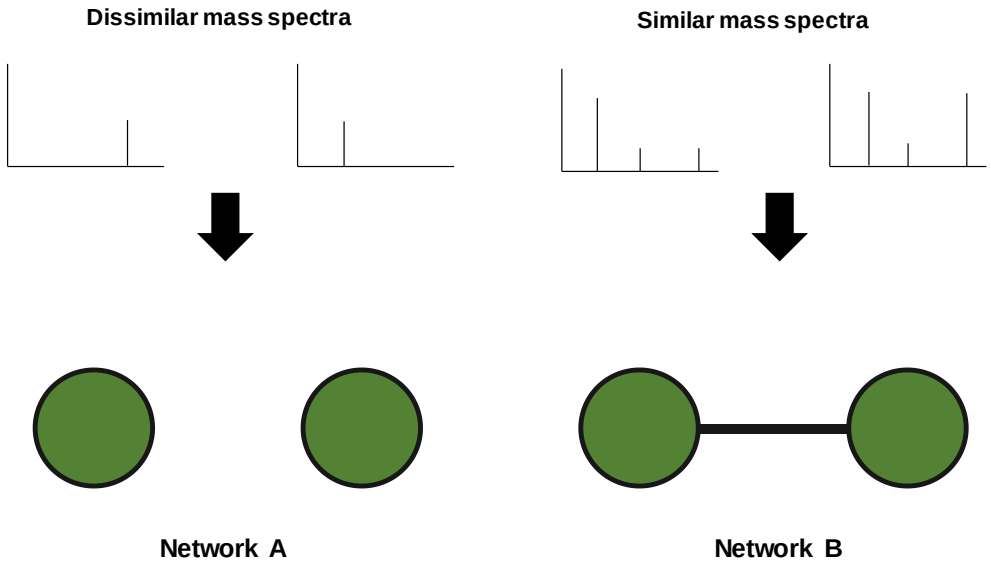


Figure 12. Representation of mass spectra as nodes in molecular networks. A) Shows two spectra that have a shared cosine score below the set cut-off point. B) Two mass spectra whose cosine score is above the cut-off point.

When comparing many spectra, those whose similarity scores are very high (usually above 0.95) are combined into consensus spectra, to reduce redundancy. Increasing the cosine score cut-off reduces the number of connections in a network, while lowering the cut-off does the opposite. A cosine score cut-off of 0.7 is the most common, however, scores as low as 0.5 are also acceptable[74].

When analyzing HRMS files by molecular networking, “seed spectra” can optionally be included, which are files of isolated compounds or standards.[69] These compounds act as starting points to which spectra from experimental files can connect if they share enough similarity. This both assists with dereplication, and has the potential to uncover new compounds with structural similarity to known ones.

After dot product calculations are completed, software such as Cytoscape can be used to map molecular networks with their nodes and edges. Networks are highly customizable and can be visually altered to highlight certain information. For example, node size can be changed to reflect the quantity of certain metabolites, or edge length may be set to represent the degree of relatedness. However, the default mapping of nodes and edges does not convey any additional information. Nodes are plotted such that the maximum number of nodes and edges are visible.

1.4 Thesis Objectives

High-resolution mass spectrometry has become a useful tool for metabolomics analyses. However, it generates large datasets that are challenging from a data processing perspective. The objective of my research is to use advanced data processing techniques including spectral molecular networking and principal component analysis to simplify the examination of high-resolution mass spectrometry data. This is done with the aim of identifying secondary metabolites from a range of fungal sources, both looking for novel compounds, and determining the chemical ecological role of known compounds. The analyses were applied to two studies involving secondary metabolite production from two different perspectives, in pathogenic and in beneficial fungi. The first analysis, detailed in Chapter 2, examines HRMS data with PCA and molecular networking to determine a metabolomic profile of pathogenic fungi that cause ginseng root rot and posits the role of these secondary metabolites in root rot disease. The second study and the focus of Chapter 3 assesses HRMS data from 302 strains of endophytic fungi isolated from Canadian fruit crops with the goal of identifying novel compound targets for isolation and characterization.

Chapter 2 – Metabolomic profiling of ginseng root rot fungi

2.1 Chapter 2 Objectives

Ilyonectria spp. cause disappearing root rot in ginseng, which heavily impacts the Canadian ginseng industry. Few secondary metabolites have been reported from these species and it is unclear if these metabolites are distinct to disease-causing isolates, or if they are ubiquitous across *Ilyonectria* spp., regardless of virulence. The goal of this study is to establish a metabolomic profile of the causative agents of ginseng root rot disease using high-resolution mass spectrometry paired with analysis by PCA and molecular networking with GNPS.

2.2 Introduction to Ginseng Root Rot

Ginseng root has been used for centuries in traditional medicines, both in Asia and among North American indigenous populations.[75] It continues to be popular, as evidenced by the $2 billion spent annually on ginseng products around the world.[76] Canada is the third-largest producer of ginseng in the world, trailing behind the gross production of both China and Korea, but nevertheless contributing significantly to the global ginseng market.[76] The ginseng species cultivated in Canada is American ginseng (*Panax quinquefolius*), which is slightly different than Asian ginseng (*Panax ginseng*).[75] Canada exports approximately 85% of its ginseng to Asia, where it is prized for its elevated levels of ginsenosides — the compounds with beneficial health effects that are found in ginseng.[75, 77, 78] The majority of ginseng grown in Canada comes from the Southwestern Ontario region, generating over $600 million annually for the Canadian economy.[79] Ginseng grows best in cool, shady plots and soil with high moisture content.[75] Unfortunately, these conditions are also perfect for the growth of ginseng root rot pathogens, which contribute to crop losses of 20-30% on average at time of harvest.[80, 81] In the controlled infections shown in Figure

13, large, soft lesions form which make the root unusable. In severe cases of root rot, the entire root rots away, or seems to disappear, thus the disease is often termed “disappearing root rot”.

Figure 13. Demonstration of healthy and diseased ginseng roots. Left: two-year-old healthy roots of American ginseng from a research plot in Simcoe, Ontario. Right: two-year-old roots from the same plot artificially infected with *Ilyonectria mors-panacis*, the main causative agent of ginseng root rot.

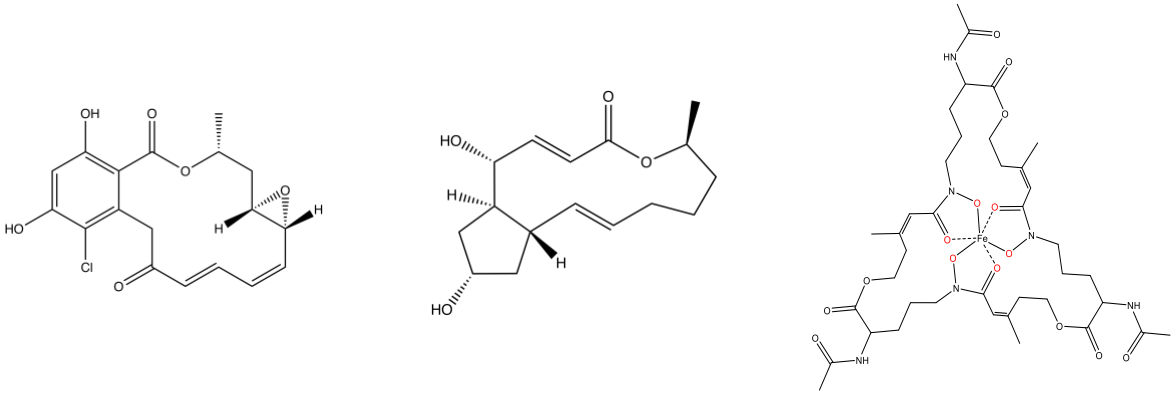
The causative agents of root rot disease are fungal pathogens from the genus *Ilyonectria*, formerly *Cylindrocarpon destructans*.[80, 82, 83] Previously, most strains isolated from ginseng root rot lesions were all categorized under the umbrella of *Cylindrocarpon destructans*. However, it was noted that not all strains described as *C. destructans* were capable of causing disease to the same degree.[80] Additionally, morphologic variation among strains cast doubt on the species assignments. Phylogenetic studies using DNA sequencing conducted over the past 20 years have reclassified many strains initially considered to be *C. destructans* as several species of *Ilyonectria* and *Neonectria*.[82, 83] The primary strains isolated from root rot lesions are typically *Ilyonectria mors-panacis* and *Ilyonectria robusta*.[84]

Because of its resilience and virulence, root rot is a contributor to ginseng replant disease—another major concern for ginseng farmers.[85] Replant disease occurs when ginseng planted in the

same soil as a previous ginseng crop fails to grow. This crop failure can persist in a field for up to 30 years, forcing farmers to rent or purchase new land for each new planting.[86] Consequently, virgin land is farther and farther from the primary growth site, which drives up production costs by increasing travel time between sites for both labourers and equipment.[86] Generous estimates give farmers 25 to 30 more years of usable land before it runs out and ginseng is no longer a viable crop in Ontario.[77]

Replant disease is believed to be a multifactorial issue, influenced by accumulation of toxic compounds in soil from ginseng plants themselves, and by changes to the soil microbiome over time.[82, 87] Root rot pathogens aggravate this issue, as they can remain in soil for years due to their hardy chlamydospores — thick-walled spores that are resistant to extreme conditions.[88]

Many studies have examined the development of root rot disease, factors that influence its progression, and morphological and genetic characteristics of fungi causing the disease. However, few studies have focused on the secondary metabolites made by these strains. Only a few metabolites have been previously identified in extracts from *Ilyonectria* and *Neonectria* spp. isolated from ginseng. These include radicicol and brefeldin A, which have antifungal and antiviral activities, respectively.[89, 90] Specifically, radicicol is potently antifungal against several species of pathogenic fungi that inhabit the same environment as ginseng pathogens. Recently, the iron-chelating siderophore N, N', N''-triacetylfusarinine C (TAFC) was also isolated from liquid media extracts of *Ilyonectria mors-panacis*.[91] This compound aids in scavenging iron from the extracellular environment and is likely important for the growth and development of *I. mors-panacis*.[91, 92]



Radicicol **Brefeldin A** **N, N', N''-Triacetylfusarinine C**

Figure 14. Structures of known metabolites isolated from cultures of the ginseng root pathogen *Ilyonectria mors-panacis*.

Apart from the compounds depicted above in Figure 14, very few other metabolites have been reported from *Ilyonectria* spp., therefore we do not have a clear picture of how or if secondary metabolites contribute to development of root rot disease. In some pathogen-host relationships, cell damage is due to cell-degrading enzymes, however, it is well-documented that some fungal pathogens cause damage by producing phytotoxic secondary metabolites. For example, *Fusarium* spp. are responsible for ear rot in corn and head blight in wheat, due to the production of deoxynivalenol (DON). DON is a mycotoxin that has toxicity in humans and livestock, and phytotoxicity to corn and wheat. It is a natural product that causes crop rotting.[93] While ginseng root pathogens may also produce phytotoxic compounds, the scarcity of compounds that are known to be produced by these pathogens, and the variety among the few compounds that have been confirmed suggests that there is a need for a greater understanding of the metabolic profiles of these species.

In this study the technique of non-targeted metabolomics was used to screen virulent and avirulent species of *Ilyonectria* and *Neonectria* for secondary metabolites. LC-HRMS data were analyzed by Principal Component Analysis to confirm that virulent species have a distinct

metabolomic profile from their avirulent counterparts. This was coupled with GNPS molecular networking of tandem HRMS data to find and identify compound classes responsible for that metabolomic profile.

2.3 Methods

2.3.1 Growth and plug extraction of fungal cultures

Twenty-two isolates of *I. mors-panacis* and other closely related species were selected from the Canadian Collection of Fungal Cultures (see Table 1, below), including one strain of *Neonectria*, included as an outgroup control. Strains were chosen that had been included in prior analyses by Seifert et al. and Cabral et al., because they have current taxonomic identifications.[80, 82] Additionally, they have information on their virulence toward ginseng root, whether by direct pathogenicity assays, or due to phylogenetic relatedness to assayed strains.[80]

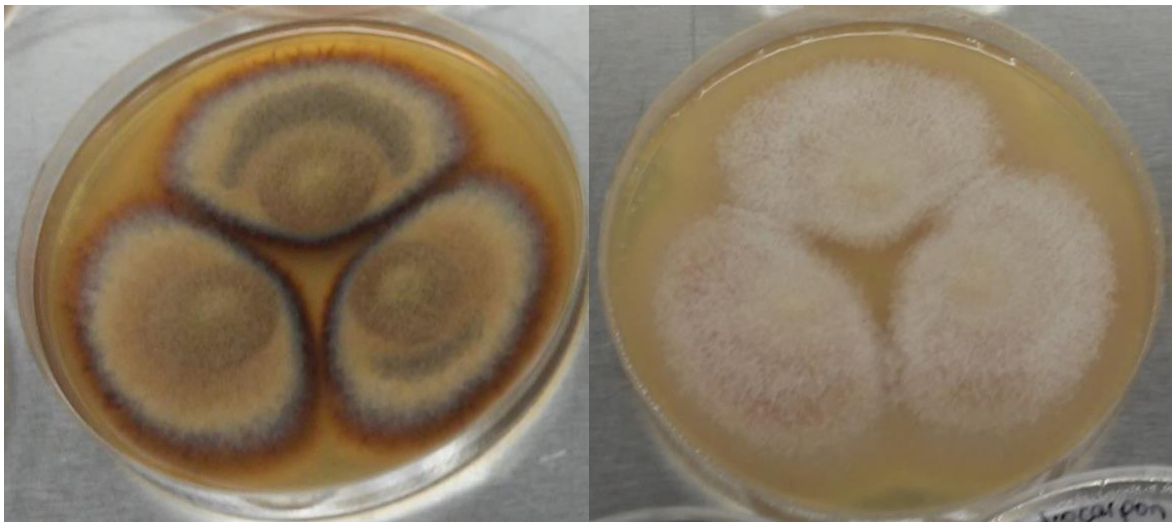


Figure 15. Three-point inoculation of *Ilyonectria* spp. on Potato Dextrose Agar. *Ilyonectria mors-panacis* (left) is capable of causing aggressive lesions on ginseng root, while *Ilyonectria rufa* (right) cannot cause lesions.

Isolates were inoculated in three points on PDA (2% (*w/v*) agar (Fisher Scientific), 2.4% (*w/v*) PDB powder (Sigma-Aldrich)) and were supplemented with 3.47 x 10^{-5} M zinc sulfate heptahydrate (Sigma-Aldrich) and 3.13 x 10^{-5} M copper sulfate (Sigma-Aldrich).[15, 94] Plates were grown in darkness at 21 °C for 20 days and extracted as per Smedsgaard et al. with modifications as follows: the number of plugs was doubled, then extracted with 4 mL of ethyl acetate, dried, then reconstituted in 1 mL of 3:1 methanol:water and filtered through a 0.45 µm PTFE filter.[9] Agar plates of *Ilyonectria* cultures may be seen in Figure 15, above. Pooled quality control (QC) samples were prepared by combining 50 µL from each extract. QC samples are included in metabolomics analyses to assess if drift is occurring in the instrument over long analysis times.[95]

Table 1. Isolates of *Ilyonectria* and *Neonectria* spp. All cultures except (*) were received from the Canadian Collection of Fungal Cultures (DAOMC).
(*) is not deposited in the DAOMC at the time of writing and is from the laboratory collection of K.A.S. (Reproduced from *Metabolites*, open access journal). [96]

ID	Species	Host	Origin
139398	*Ilyonectria robusta*	*Prunus cerasus* (Sour cherry)	Canada, ON
144524	*I. torresensis*	*Vitis vinifera* (Grape)	ON
220159	*I. mors-panacis*	*Panax quinquefolius* (American ginseng)	ON
226721	*I. rufa*	*Pseudotsuga menziesii* (Douglas fir)	Canada, BC
226727	*I. mors-panacis*	*P. quinquefolius*	ON
226729	*I. robusta*	*P. quinquefolius*	ON
226730	*I. estremocensis*	*Picea glauca* (White spruce)	Canada, QC
230337	*I. mors-panacis*	*Panax ginseng* (Asian ginseng)	Japan
230338	*I. mors-panacis*	*P. ginseng*	Japan
234582	*I. mors-panacis*	*P. quinquefolius*	ON
251601	*I. mors-panacis*	*P. quinquefolius*	ON
251602	*I. mors-panacis*	*P. quinquefolius*	ON
251603	*I. mors-panacis*	*P. quinquefolius*	ON
251604	*I. mors-panacis*	*P. quinquefolius*	ON
251605	*I. mors-panacis*	*P. quinquefolius*	ON
251606	*I. mors-panacis*	*P. quinquefolius*	ON
251607	*I. mors-panacis*	*P. quinquefolius*	ON
251608	*I. rufa*	*P. menziesii*	BC
251609	*I. rufa*	*P. glauca*	QC
251610	*I. mors-panacis*	*P. quinquefolius*	ON
251611	*I. mors-panacis*	*P. quinquefolius*	ON
*94-1356	*Neonectria obtusisporum*	*Picea mariana* (Black spruce)	QC

2.3.2 LC-MS and LC-MS/MS analysis of plug extracts

High resolution MS/MS analysis of secondary metabolites was performed using a Q-Exactive Orbitrap mass spectrometer (Thermo Fisher Scientific) with a heated electrospray ionization (HESI) source connected to an Agilent 1290 UHPLC system. Positive and negative ionization modes were assessed for both HRMS and tandem HRMS. Separation was done with a dual-solvent system with water + 0.1% (*v/v*) formic acid and acetonitrile + 0.1% (*v/v*) formic acid

(solvents A and B, respectively), at a flow rate of 0.3 mL/min. The gradient was held at 0% B for 0.5 minutes, increased to 100% B over 3 minutes, held at 100% B for 2.5 minutes, decreased to 0% B over 0.5 minutes and finally held at 0% B for one minute. For each sample, 5 µL was injected onto an EclipsePlus RRHD C-18 column (2.1x50 mm, 1.8 µm; Agilent) heated to 35°C. HESI conditions were as follows: capillary temperature, 400°C; sheath gas, 19 units (positive mode) and 17 units (negative mode); auxiliary gas, 8 units; probe heater temperature, 450°C; S-Lens RF level, 45; capillary voltage, 3.9 kV (positive mode) and 4.0 kV (negative mode).

All samples were analyzed by HRMS with the following settings: resolution, 140,000; automatic gain control (AGC) target, 5 x 10_5; max IT, 512 ms; scan range, 80-1200 *m/z*. QC samples were injected at the beginning, end, and periodically throughout analysis[95]. Samples were also analyzed by tandem HRMS, with the following settings: resolution, 17,500; AGC target, 1 x 10_6; top N, 10; isolation window, 1.2 *m/z*; intensity threshold, 5.5 x 10_5; dynamic exclusion, 5.0 s; max IT, 64 ms; scan range, 200-2000 Da; NCE, 35.

2.3.3 Principal Component Analysis

For Principle Component Analysis (PCA), Thermo Raw HRMS files were converted into centroid mode using MSConvert and filtered with the peak picking parameters set at MS level 1 (Vendor algorithm).[97] This returned mzML files that are usable in R.[98] The mzML files were processed in R (3.5.3) with the software package *xcms* (3.2.0).[38, 99, 100] Parameters used in *xcms* processing were as follows: method="centWave", prefilter=c(5, 5000), ppm=5, snthresh=10, peakwidth=c(5,30), noise=1000000, bw=5, minfrac=0.001, mzwid=0.015. All zero values were imputed with two-thirds of the lowest value measured per metabolite so PCA could be performed.[101] Metabolites that were only found in blank media were excluded, and isotopic peaks were removed manually. Peak area values were log10-transformed to correct for heteroscedasticity

in the data.[102] The data were also pareto scaled to account for large fold differences, then PCA was performed with the R packages *FactoMineR* (2.3) and *MetabolAnalyze* (1.3.1).[102-105] Additionally, factor loadings values were calculated with the same packages listed above. The first and second principal component scores were plotted against each other. Positive and negative ionization modes were analyzed separately. A Kruskal-Wallis test for non-parametric data was performed on metabolite peak areas from Groups 1 and 2 using R, and Benjamini-Hochberg *p*-value correction was applied.[63] Metabolites with a corrected p-value < 0.05 were considered to be significantly different between groups.

2.3.4 Molecular Networking Parameters and Visualization

For molecular networking in GNPS, Thermo Raw HRMS/MS files analyzed in positive ionization mode were converted to mzML format with MSConvert[97] with the following settings: peak picking from MS level 1-2, 32-bit binary encoding precision, and no file compression, as per recommendations from GNPS. Files were uploaded to GNPS with FileZilla (3.9.0.5) and analyzed with the following parameters: precursor ion mass tolerance, 0.001 Da; fragment ion mass tolerance, 0.001 Da; min pairs cosine score, 0.6; Network TopK, 10; maximum connected component size, 100; minimum matched fragment ions, 4; minimum cluster size, 2; MSCluster, on. Three in-house MS/MS files of radicicol and two other unknown compounds purified from extracts of *I. mors-panacis* strain DAOMC 251601 were included as seed spectra. The unknown compounds have chemical formulas of $C_{18}H_{20}O_5$ and $C_{18}H_{20}O_6$. The GNPS output was downloaded as a GRAPHML file and visualized in Cytoscape 3.6.1.[106] To simplify, all features that were only present in the blank media and clusters comprised of only seed spectra were removed from the network. Isotopic peaks and peaks with poor chromatography were removed manually.

2.4 Results and Discussion

2.4.1 Results and Discussion of Principal Component Analysis

LC-HRMS data were processed in R, using the package *xcms*. After removing features attributed to media, 1813 and 1019 features were detected in positive and negative ionization modes, respectively. PCA was performed in R, using the features lists that were output from *xcms*, which contain metabolite peak areas acquired by LC-HRMS. Plots were then generated for both positive and negative ionization modes that represent variance between samples. The number in parentheses on each axis describes the percentage of total variance that the principal component accounts for. Sample plots are depicted in Figure 16 A and C, where each point represents a fungal strain. Factor loadings for samples are pictured in Figure 16 B and D, below. Each ray represents an LC-MS feature and its orientation indicates which group or sample has the highest abundance of that feature. Rays highlighted in red were significantly different between Groups 1 and 2, according to a Kruskal-Wallis test for non-parametric data. Batch effect was considered to be minimal, because QC samples included in initial versions of plots grouped closely together (see B1, Appendix B).

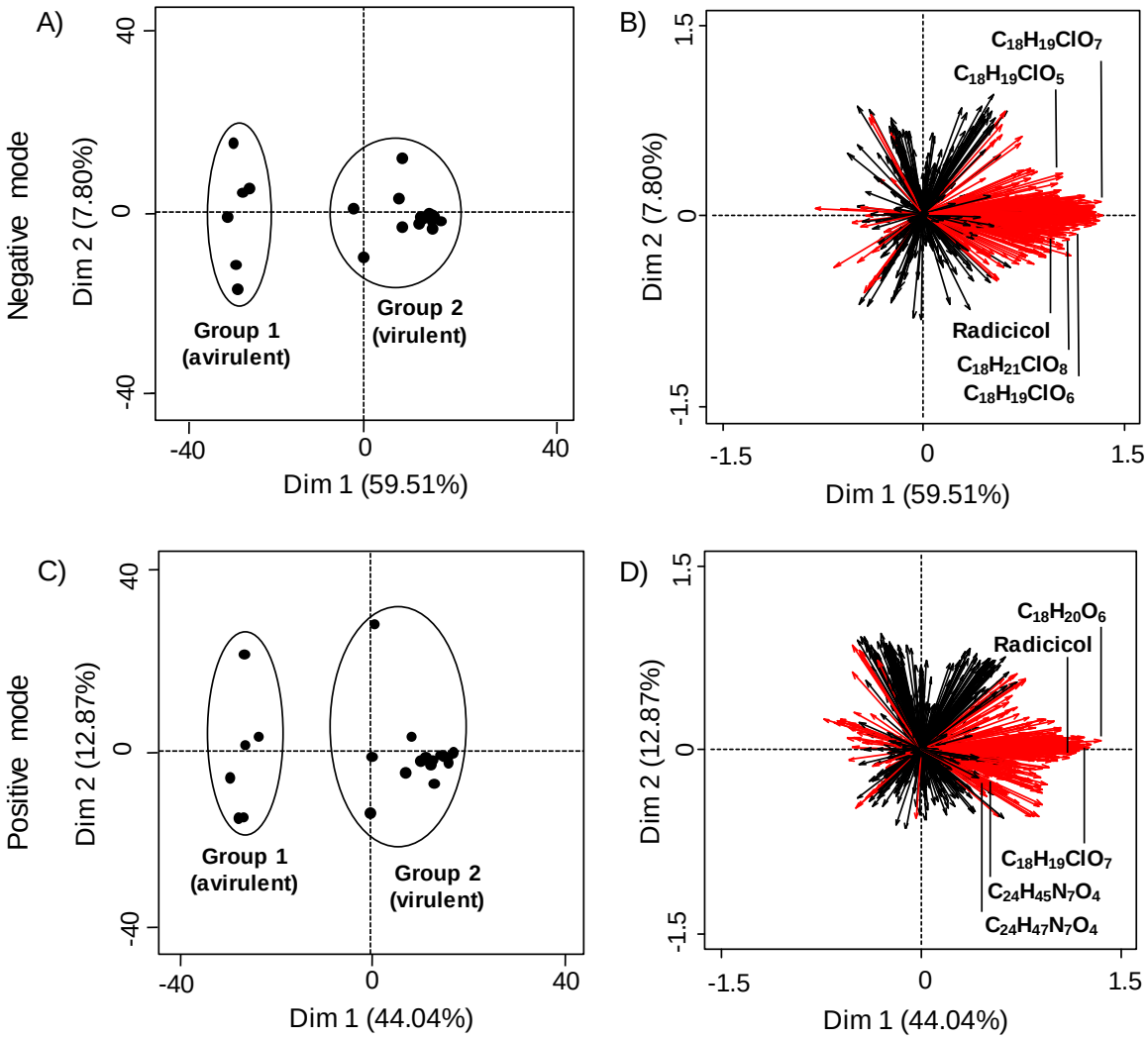


Figure 16. Principal Component Analysis plots based on metabolites from *Ilyonectria* and *Neonectria* spp. Plots A and C show individual samples. Plots B and D show factor loadings for samples, with each ray representing an LC-MS feature. Red rays indicate features with $p < 0.05$ (Kruskal-Wallis test with Benjamini-Hochberg *p*-value correction). Black rays were not statistically significant after BH correction. The five most abundant metabolites in each ionization mode are indicated. (Reproduced from *Metabolites*, open access journal).[96]

In both positive and negative ionization modes, samples divided into two distinct groups, which are listed in Table 2, below.

Table 2. Grouping of *Ilyonectria* and *Neonectria* strains by Principal Component Analysis.

Group 1 (avirulent strains)	Group 2 (virulent strains)		
DAOMC 144524	DAOMC 139398	DAOMC 234582	DAOMC 251606
DAOMC 226721	DAOMC 221059	DAOMC 251601	DAOMC 251607
DAOMC 226730	DAOMC 226727	DAOMC 251602	DAOMC 251610
DAOMC 251608	DAOMC 226729	DAOMC 251603	DAOMC 251611
DAOMC 251609	DAOMC 230337	DAOMC 251604	
Strain 94-1356	DAOMC 230338	DAOMC 251605	

Interestingly, DAOMC 230337 and 230338 are isolates of *I. mors-panacis* isolated from Asian ginseng in Japan but they share a similar chemical profile to isolates from Ontario. This mirrors the observation from Seifert et al. that these same strains of *I. mors-panacis* have a high degree of genetic similarity, despite having been recovered from different geographic regions and ginseng species.[80]

Furthermore, both strains of *I. robusta* share a chemical profile with strains of *I. mors-panacis*, consistent with reports that *I. robusta* is also capable of causing root rot, and perhaps indicating that secondary metabolites are important in virulence toward ginseng root.[84] Overall, the PCA plots demonstrate that in the growth conditions of this experiment, aggressive strains of *Ilyonectria* make a higher abundance and wider variety of compounds than weakly aggressive strains do. Upon manual inspection of mass spectra, nor brefeldin A, nor TAFC were present in any sample. It is likely that growth conditions in this experiment did not favour their production. Cultures in this study were grown on solid media, whereas brefeldin A and TAFC have only been isolated from liquid culture.[107]

To determine which metabolites are responsible for the division of groups, a Kruskal-Wallis test was performed. A total of 416 and 418 features were significantly different ($p < 0.05$ after Benjamini-Hochberg correction) in negative and positive ionization modes, respectively.

Most statistically significant metabolites were produced solely by virulent strains. The five most abundant compounds made by Group 2 were assessed in both negative and positive ionization modes. These are presented in Table 3, below, and are highlighted in Figure 16 B and D, above. Additional abundant compounds in positive and negative ionization modes are listed in Appendix B. The most abundant compound in both ionization modes was radicicol, confirmed by comparing to experimental ESI-Q-TOF MS/MS data retrieved from METLIN.[89, 108] Several compounds have chemical formulas similar to that of radicicol. To determine if these compounds were structurally similar to radicicol, MS/MS data from these samples were subjected to spectral molecular networking via GNPS. No compounds made solely by avirulent strains of *Ilyonectria* were identifiable under these experimental conditions.

Table 3. The five most abundant compounds made by *Ilyonectria* spp. *m/z* values and chemical formulas from HRMS data are shown for positive and negative ionization modes. (*) indicates radicicol (confirmed by MS/MS) which was the most abundant compound in both ionization modes. Additional compounds are listed in Appendix B. (Reproduced from *Metabolites*, open access journal).[96]

Positive		**Negative**	
m/z	**Formula**	***m/z***	**Formula**
355.1150 $[M+Na]^+$	$C_{18}H_{20}O_6$	349.0847 $[M-H]^-$	$C_{18}H_{19}ClO_5$
365.0785* $[M+H]^+$	$C_{18}H_{17}ClO_6$	363.0641* $[M-H]^-$	$C_{18}H_{17}ClO_6$
383.0890 $[M+H]^+$	$C_{18}H_{19}ClO_7$	365.0795 $[M-H]^-$	$C_{18}H_{19}ClO_6$
496.3631 $[M+H]^+$	$C_{28}H_{49}NO_6$	381.0746 $[M-H]^-$	$C_{18}H_{19}ClO_7$
498.3787 $[M+H]^+$	$C_{28}H_{51}NO_6$	399.0850 $[M-H]^-$	$C_{18}H_{21}ClO_8$

The list of the most abundant compounds was slightly different in positive mode compared with negative mode. This is likely due to some compounds ionizing preferentially in one mode over the other. Despite the difference, PCA plots demonstrate that virulent species of *Ilyonectria* have a distinct chemical profile from avirulent species.

2.4.2 Results and Discussion of Molecular Networking

While PCA is useful in metabolomics for determining similarities or differences among samples, it is difficult to rely solely on PCA for compound identification. Feature lists generated by *xcms* can comprise thousands of compounds, and dereplication is especially challenging for lists containing only *m/z* values and retention times. Molecular networking supplements PCA by guiding us to toward classes of compounds and giving by visual targets to begin working from. The molecular network displayed in Figure 17, below, was generated with positive mode MS/MS data, as seed spectra were only available in positive ionization mode. A network was also made with negative mode data, however the lack of seed spectra caused it to be of poor quality, indicated by a smaller number of nodes and few connections compared with the positive mode network (see Appendix B, Figure B2). Therefore, it was not explored further.

The five most abundant compounds from Table 3 are divided among three clusters, labeled A-C. Cluster A contains radicicol (from seed spectra and from samples) as well as $C_{18}H_{19}ClO_7$, from Table 3. Cluster B contains the compounds $C_{28}H_{49}NO_6$ and $C_{28}H_{51}NO_6$. After colouring nodes according to the group they came from, it became clear that these two compounds are produced by both virulent and avirulent strains. Although the Kruskal-Wallis test is useful for discerning which metabolites are different between groups, it cannot determine whether metabolites are unique to a sample group. Because neither of the nitrogen-containing compounds were dereplicable under these experimental conditions, and because they were not unique to virulent species, they were not pursued for further identification.

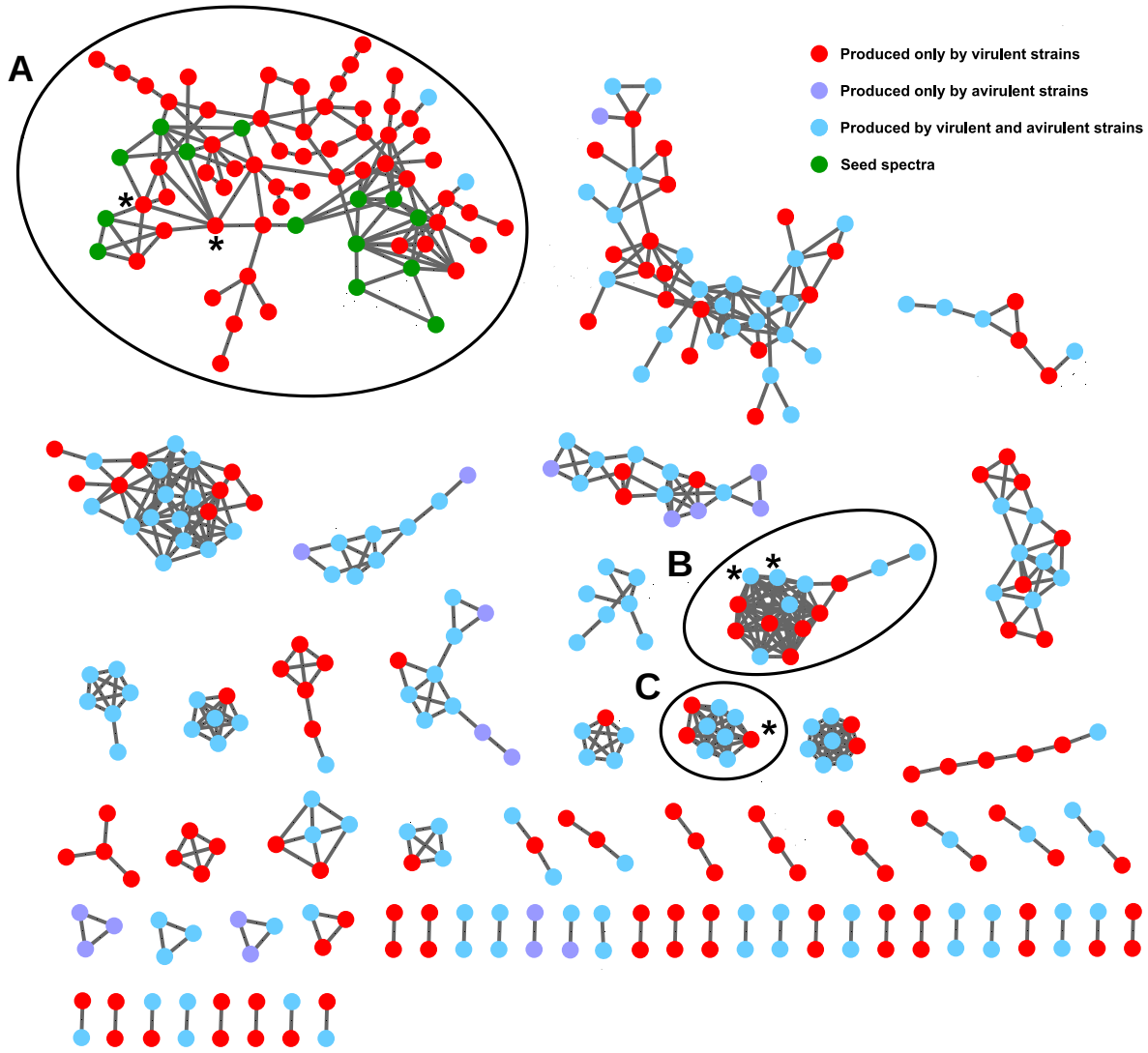


Figure 17. GNPS molecular network of positive mode MS/MS data from strains of *Ilyonectria* and *Neonectria* visualized in Cytoscape. Red: features made only by aggressive strains of *Ilyonectria*. Violet: made only by weakly aggressive strains of *Ilyonectria* and *Neonectria*. Blue: made by both types of strains. Dark green: features from seed spectra. The five compounds with the highest relative abundance are indicated by asterisks. (Reproduced from *Metabolites*, open access journal).[96]

Finally, in cluster C, the remaining *m/z* value from Table 3 was determined to be the sodiated ion of a metabolite found within cluster A. Because cluster A is comprised almost exclusively of features from virulent strains, it was examined more closely for virulent-specific metabolites (Figure 18, below).

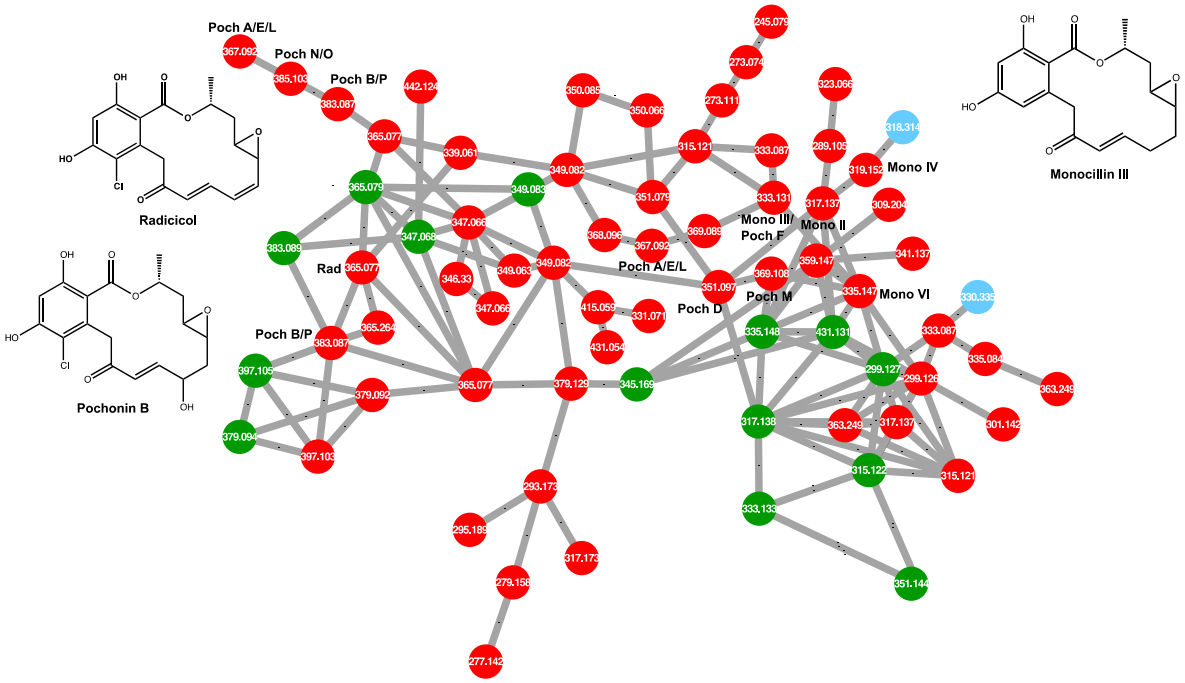


Figure 18. Close-up of cluster A from GNPS molecular network of positive mode ESI-HRMS data from strains of *Ilyonectria* and *Neonectria* spp. Red: features produced only by virulent species of *Ilyonectria*. Green: features from seed spectra. Blue: features produced by both virulent and avirulent species. Structures are of compounds from Table 3. (Poch—pochonin; Mono—monocillin; Rad—radicicol). (Reproduced from *Metabolites*, open access journal).[96]

Cluster A includes several chlorine-containing compounds with formulas similar to that of radicicol that are connected by edges within the network, indicating they share spectral similarity. From Table 3, the compounds with formulas of $C_{18}H_{20}O_6$ and $C_{18}H_{19}ClO_7$ match the formulas of monocillin III and pochonin B, respectively. These belong to a family of compounds named resorcyclic acid lactones (RALs), which also includes radicicol. The cluster also contains ten other nodes that have chemical formulas matching those of several other RALs. These include nodes putatively identified as pochonins A, B, D-F, L-P, and monocillins II-IV and VI. Some nodes are labeled with more than one putative identification because there are several isomeric RALs that

were not distinguishable in these conditions. Structures of putatively identified compounds are shown in Figure 19, below.

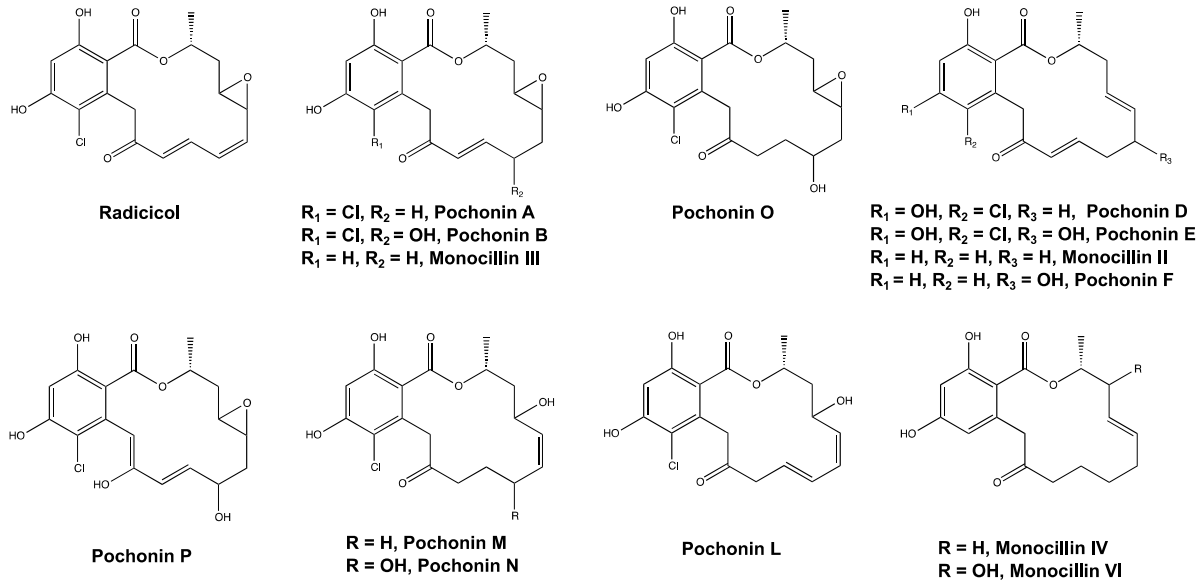


Figure 19. Structures of compounds putatively identified from ethyl acetate extracts of *Ilyonectria mors-panacis* and *Ilyonectria robusta*. (Reproduced from *Metabolites*, open access journal).[96]

Radicicol has been observed in conjunction with pochonins in other species within the same order as *Ilyonectria*, namely *Pochonia chlamydosporia* and *Monocillium nordinii*[109, 110] but not previously in *Ilyonectria* spp. The observation of RALs within the same taxonomic order as *Ilyonectria* is somewhat expected, as the biosynthetic machinery used for certain natural product pathways is often conserved evolutionarily. While several chemical formulas predicted from the five most abundant *m/z* values were initially suggestive of pochonins, their connection to radicicol within the molecular network provides confidence in their identification.

Several RALs have biological activities, including antiviral and antiparasitic activities.[109-111] The mechanism of action of these compounds is likely through a combination of biological

pathways. Studies by Moulin et al. have shown that pochonins A and D are inhibitors for heat shock protein 90 (HSP90), as is radicicol.[112, 113] HSP90 is a ubiquitously expressed chaperone protein that assists in protein folding for many proteins within the cell.[114] Among many other activities, HSP90 is involved in regulating chitin synthase, which builds cell walls in fungi.[115] When disrupted, chitin synthase is unable to repair or build cell walls. Antimicrobial RALs may indirectly assist ginseng root pathogens to colonize their hosts by interrupting the growth of other environmental fungal pathogens. These compounds may be produced to help defend *Ilyonectria* spp. against other pathogens, and as a result, they have an improved chance of survival and of infecting ginseng roots.

PCA and molecular networking with GNPS permit rapid analysis of complex mixtures and allow for tentative compound identification. However, these tools are somewhat limited in their ability to confirm compound identities without the inclusion of standards. Putatively identified compounds must be isolated by HPLC and characterized by NMR to confirm their identities. Although spectral molecular networking does have this drawback pertaining to individual compounds, it is nonetheless a powerful tool for identifying compound classes and for simplifying the analysis of metabolomics data. In this application, spectral molecular networking identified a class of compounds that are specific to virulent strains of *Ilyonectria*. PCA differentiated the metabolomic profiles of virulent and avirulent strains of *Ilyonectria* and *Neonectria*. Although there are still many nodes that may be relevant to root rot in the overall network, the identification of RALs within the central radicicol-containing cluster provides a starting point for future analysis.

2.5 Conclusions and Suggestions for Future Work

PCA determined that species of *Ilyonectria* capable of infecting ginseng root have a distinct chemical profile from species that are not capable of infecting ginseng. *I. mors-panacis* and *I. robusta* made a high abundance of radicicol and several other related compounds. Spectral molecular networking with GNPS confirmed a class of compounds related to radicicol, putatively identified as RALs, that have not been previously reported from *Ilyonectria* spp. The results of PCA and molecular networking will guide the isolation and characterization of both known and potentially novel structural analogues and derivatives to RALs that can be used in future toxicity tests.

While several antifungal compounds have been identified in plug extracts of *Ilyonectria mors-panacis* on PDA, there were no identifiable phytotoxic compounds. It may be that *I. mors-panacis* does not produce phytotoxic compounds, or it may be that the medium these samples were grown on does not contain the proper growth factors and secondary messengers to promote the production of compounds it would make on ginseng roots. Further experiments ought to focus on preliminary phytotoxicity tests that may be completed with crude extracts and isolated compounds from *I. mors-panacis* and *I. robusta* to determine if they are capable of damaging plant material. One route is to test extracts using a model organism of toxicity such as duckweed (*Lemna minor*), which is a plant that often used to test toxicity of wastewater effluent, and occasionally mycotoxins because of its sensitivity to environmental changes.[116-118] Alternately, if fresh ginseng leaves are available, an experiment similar to that described by Cossette and Miller may be accomplished to determine if electrolyte loss occurs in host leaves due to the presence of *Ilyonectria* extracts. Electrolyte leakage is a quantifiable measure of cell damage.[93]

Additionally, determining what types of compounds *I. mors-panacis* makes when grown on ginseng root is pertinent, especially at the host-pathogen interface. This may be accomplished by infecting ginseng roots with *I. mors-panacis* and allowing lesion formation, then extracting lesions to analyze by HRMS, or by cross-sectioning roots and analyzing the lesions by mass spec imaging techniques such as desorption electrospray ionization (DESI) or matrix-assisted laser desorption ionization (MALDI). Knowing which compounds are produced both within lesions and where the pathogen and host interact may provide answers about the importance of secondary metabolites to lesion formation.

Chapter 3: Non-targeted screening of natural products from 302 fungal endophytes isolated from Canadian fruit crops

3.1 Chapter 3 Objectives

Fungal endophytes are a rich and underexplored source of natural products. Canadian fruit crops are hosts to many diverse and novel endophytes and often receive protection from the natural products their symbionts produce. It is of interest to screen endophytes for new compounds, because many of them produce bioactive natural products with uses in agriculture, pharmacology, and in chemistry research. However, traditional screening methods for compound discovery can be lengthy and difficult processes. Therefore, the objective of this study is to screen a library of 302 extracts from fungal endophytes isolated from Canadian fruit crops using spectral molecular networking with GNPS to identify new or novel natural product targets that will be isolated and characterized in further studies.

3.2 Introduction

While many fungal species are pathogenic and cause diseases that negatively impact the agricultural sector, there are also species that live symbiotically with plants and provide beneficial effects to their host. Endophytic fungi are fairly ubiquitous and reside inside the tissues of many plants (Figure 20), and generally provide them with a survival advantage, such as increased resistance to pests or pathogens.[119] In return for the protection they provide, endophytes receive nutrition from their host. There are several examples of endophytes being used in agricultural products with great success. Consumers may purchase grass seed treated with endophytic fungi to protect it against insect damage.[120] The Canadian lumber company Irving inoculates their saplings

with an endophyte native to spruce trees that protects against spruce budworm, saving saplings from succumbing to the pest.[121]

While endophytes may be exploited directly as in the examples above, another area of interest is in screening these relatively understudied fungal species for new or novel natural products that may be used as pesticides, pharmaceuticals, or in research. Generally, the protective advantages afforded by endophytes to their hosts are due to unique secondary metabolites they produce. In the examples given above, grass endophytes produce toxic ergot alkaloids that ward off pests, while the endophyte used by Irving produces high levels of the anti-insectan compound rugulosin.[122] By examining the natural products of lesser-known fungi, one may discover compounds that can be used in medicine or agriculture, or that answer questions about known biological relationships.

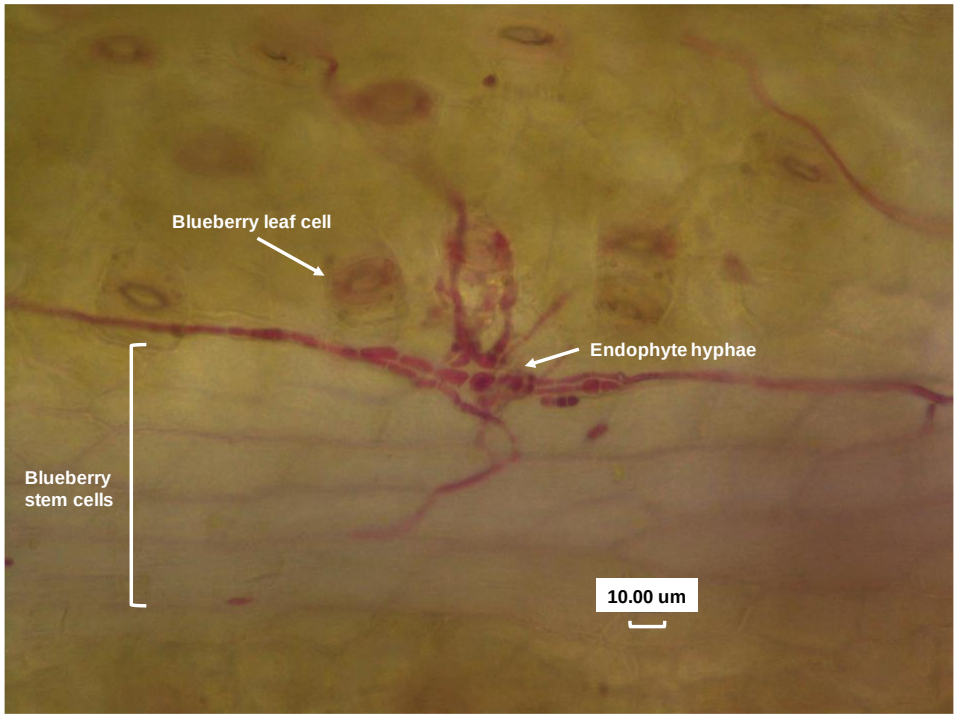


Figure 20. Microscope image of a blueberry leaf infected with an unidentified fungal endophyte, at 400 x magnification. Endophyte cells are stained with Rose Bengal and can be seen as pink tendrils.

Many new and novel compounds have been reported as natural products of endophytes in recent years, often with bioactivity. Endophytes living symbiotically with Canadian fruit crops such as blueberries and grapes have proven to make an abundance of natural products with interesting bioactivities. These include the recently published ellisiiamides from *Xylaria ellisii* — an endophyte isolated from blueberry and pine plants on Canada's east coast.[123, 124] These newly discovered compounds are cyclic pentapeptides which are posited to have anti-insectan activity, based on their structural relatedness to other known anti-insectan compounds.[123] Unique trienylfuranones have also been discovered from *Hypoxylon submonticulosum* — a raspberry endophyte that was isolated from plants from Jordan Station, Ontario. These have biological activity against *Saccharomyces cerevisiae*, which is a model fungal organism.[125] Nemanilactones and nemanifuranones discovered from the grape endophyte *Nemania serpens* have biological activity against several strains of bacteria and yeast.[126] Structures of the aforementioned compounds are depicted below, in Figure 21.

While endophytes are interesting to study for their natural products, the methods traditionally used to screen large numbers of fungal isolates are not always efficient. Fungi are grown in liquid culture to acquire enough crude extract to perform filter paper assays. Crude extract is applied to paper filter discs which sit on agar plates of microbial cultures—usually a few different species, including Gram-negative and -positive bacteria. Endophytic extracts that affect the growth of these select organisms must then be grown in large-scale fermentations to achieve a large enough amount of material to isolate by HPLC and analyze by NMR spectroscopy. There are several drawbacks to this method, however. The number of organisms that are tested in bioassays is usually limited to two or three different species. Therefore, compounds that are not bioactive against tested organisms but that do have activity against untested organisms would be

overlooked. Additionally, there are thousands of bioactive compounds that have already been discovered but it can be difficult to determine which ones are present in an extract without the inclusion of standards. It is not feasible for most studies to include standards of thousands of known compounds with their unknowns with which to compare. Thus, spectral molecular networking was chosen as a method to identify new or novel compounds from the collection of endophytes, because it highlights groups of structurally related compounds, and also permits dereplication of some previously identified compounds when screened with the GNPS spectral library.

Interestingly, although the overall number of new natural products discovered each year is increasing, the number of products with unique structural classes is not increasing.[127] Many newly discovered compounds are slight variations on known structures, as seen in Figure 21. This is likely because the process of making secondary metabolites generally requires many biosynthetic steps, any intermediate of which could have its own distinct function.[128] Luckily, this is advantageous for many of our natural product needs. A pressing issue in healthcare is the evolution of antibiotic-resistant strains of deadly infectious agents. The pharmaceutical industry is in a constant race against the evolution of these strains to find antibiotics that are effective against them. By searching for compounds that share structural similarity to known bioactive compounds, one may find structures that are lethal to pathogens of concern.

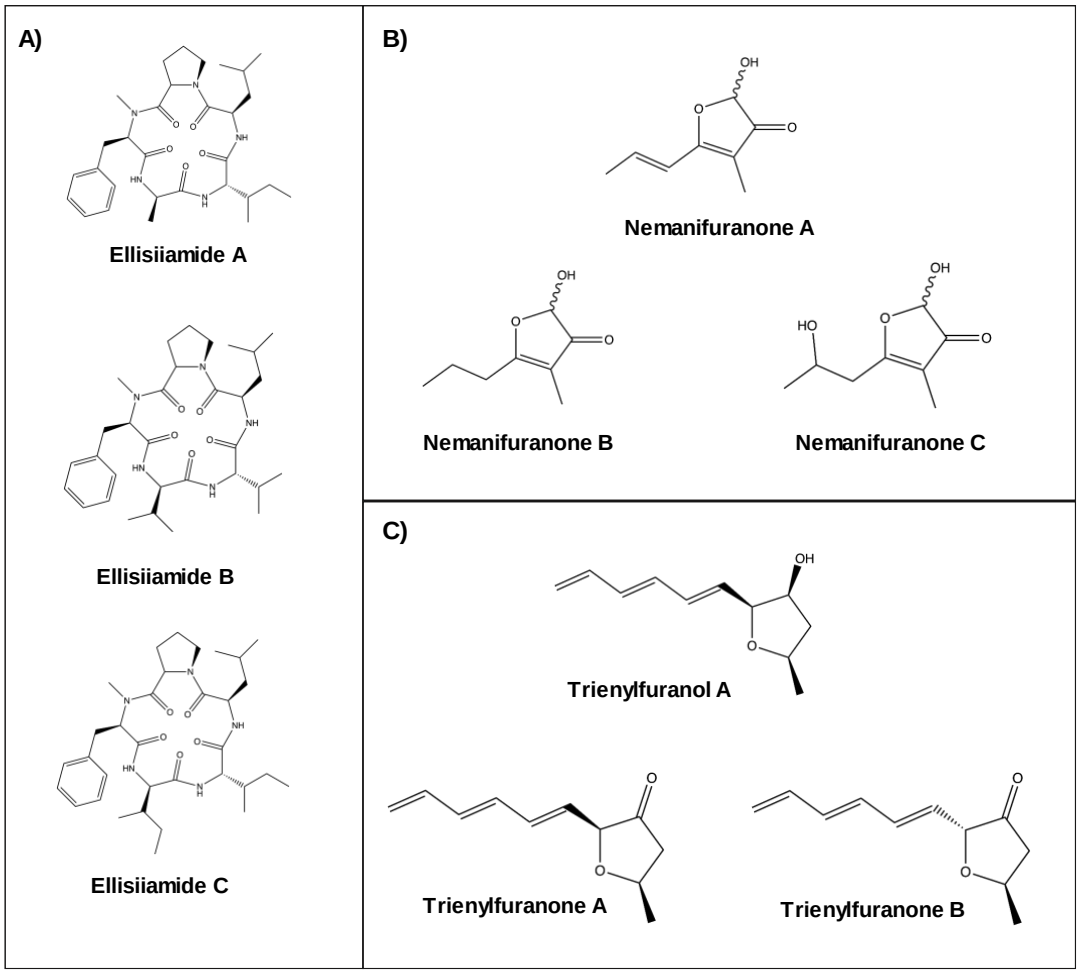


Figure 21. Select bioactive natural products from fungal endophytes. A) Cyclic pentapeptides from *Xylaria ellisii* with suspected anti-insectan properties. B) Antimicrobial furan compounds from *Nemania serpens*. C) Furan-type compounds from *Hypoxylon submonticulosum* with antifungal activity.

Seed spectra that were included in the molecular networking analysis are solutions that contain known compounds and were previously isolated from fungal endophytes, including *Xylaria ellisii*, *Coniochaeta tetraspora*, and *Xylaria castorea*. This study used non-targeted metabolomics to screen a group of 302 fungal endophytes for new and novel natural products. LC-HRMS data were analyzed by GNPS to find several new compounds that are related to known bioactive compounds, making them good targets to isolate and characterize in future studies.

3.3 Methods

3.3.1. LC-HRMS analysis of endophyte cultures

Endophytes used in this study were isolated from Canadian fruit crops between 2011 and 2015 using the method outlined by Ginn (1998).[129] Briefly, leaves and stems from blueberries, cranberries, raspberries, and grapes were surface sterilized with bleach and ethanol, then samples were sliced into pieces and placed on agar plates, exposing the interior to the nutrient-rich environment. If endophytes are present, they grow out of the host plant tissue towards the medium where they can be isolated and identified. DNA was extracted and sequenced using internal transcribed spacer gene (ITS) primers over the same time period by Megan Kelman (AAFC), and DNA sequences were interpreted via BLAST by Dr. Joey Tanney (NRCAN, Pacific Forestry Centre, Victoria, BC). ITS is considered the universal fungal DNA barcode because its sequence is usually specific to a species, therefore it is frequently the first gene examined for fungal identification.[130] Endophyte extracts were grown and extracted over the course of 2016. Endophytes were grown on 20 mL of PDA at 23 °C in the dark for 2-6 weeks, or until cultures reached confluence. The entire plate was extracted by homogenizing the agar and cells with 20 mL of methanol, gravity filtering the extract with a #1 Whatman filter and taking a 1-mL aliquot of the filtrate. Aliquots were dried by nitrogen evaporation and dried samples were stored at -20 °C until analysis.

Dried endophyte extracts and isolated compounds used as seed spectra were re-suspended in 1 mL of methanol (Fisher Scientific, Fair Lawn, NJ, USA) and analyzed by LC-HESI-HRMS on a Thermo Q-Exactive Orbitrap mass spectrometer coupled with an Agilent 1290 UHPLC system. A list of all endophytes, their species assignments, and isolation information is available in Appendix C. The list of isolated compounds that were used as seed spectra is detailed in Table

4, below. Pooled quality control (QC) samples were produced by combining 10-uL aliquots from each sample.

Table 4. Seed spectra used in molecular network of Canadian fungal endophytes. Seed spectra included in the network are bioactive compounds previously isolated from fungal endophytes.

Name	Calc *m/z* $[M+H]_+$	Formula	Fungal source
Coriloxin	171.0652	$C_8H_{10}O_4$	*Xylaria castorea*
Abscisic acid	265.1434	$C_{15}H_{20}O_4$	*Nigrospora sphaerica*
Ascochitine	277.1071	$C_{15}H_{16}O_5$	*Coniochaeta tetraspora*
7-hydroxy-3-(hydroxymethyl)-2-(2-hydroxypropyl)-6-methoxy-4H-chromen-4-one (fulvic acid derivative)	281.1019	$C_{14}H_{16}O_6$	*Sphaerulina rhabdoclinis*
4,10-Dihydro-3,7,8-trihydroxy-3-methyl-10-oxo-1H,3H-pyrano[4,3-b][1]benzopyran-9-carboxylic acid (fulvic acid analogue)	309.0605	$C_{14}H_{12}O_8$	*Sphaerulina rhabdoclinis*
Zygosporin E	492.2744	$C_{30}H_{37}NO_5$	*Xylaria ellisii*
Cytochalasin D	508.2693	$C_{30}H_{37}NO_6$	*Xylaria ellisii*
Epoxycytochalasin D	524.2642	$C_{30}H_{37}NO_7$	*Xylaria ellisii*
Ellisiiamide A	556.3493	$C_{30}H_{45}N_5O_5$	*Xylaria ellisii*
Ellisiiamide B	570.3650	$C_{31}H_{47}N_5O_5$	*Xylaria ellisii*
Cyclic pentapeptide 1	584.3806	$C_{32}H_{49}N_5O_5$	*Xylaria ellisii*
Ellisiiamide C	598.3963	$C_{33}H_{51}N_5O_5$	*Xylaria ellisii*
Ellisiiamide G	600.3755	$C_{32}H_{49}N_5O_6$	*Xylaria ellisii*
Hirsutatin A	677.3756	$C_{34}H_{52}N_4O_{10}$	*Xylaria ellisii*

Chromatographic separation was completed with a dual-solvent system with acetonitrile + 0.1% formic acid (solvent A), and water + 0.1% formic acid (solvent B) at a flow rate of 0.3 mL/min. The gradient was held at 0% B for 0.5 minutes, increased to 100% B over 3 minutes, held at 100% B for 2.5 minutes, then decreased to 0% B over 0.5 minutes and held at 0% B for 1 minute. All samples had 5 µL injected on an EclipsePlus RRHD C-18 column (2.1x50 mm, 1.8 µm;

Agilent) maintained at 35°C. HESI conditions were as follows: capillary temperature, 400°C; sheath gas, 17 units; auxiliary gas, 8 units; probe heater temperature, 450°C; S-Lens RF level, 50; capillary voltage, 3.9 kV.

Data were acquired in positive ionization mode with data-dependent acquisition using the following settings: resolution, 70,000; automatic gain control (AGC) target, 1×10^6; max IT, 256 ms; scan range, 100-1500 *m/z*. The 10 ions with highest intensity were selected from each MS scan to be fragmented by MS/MS at resolution 17,500; AGC target 1×10^6; max IT, 64 ms; stepped NCE, 28/50; isolation window, 1.2 *m/z*; intensity threshold, 1.3×10^5; dynamic exclusion, 10.0 s. QC samples were injected at the beginning, end, and periodically throughout the run to assess instrument drift.[95]

3.3.2 Data processing and analysis by PCA

Thermo Raw data files were converted to mzML format using MSConvert with the following settings: 32-bit binary encoding precision, no file compression, and peak picking from levels 1-2.[97] This produced files that are usable both with R statistical software, and later on, with GNPS. The mzML files were brought into R (3.5.3) to perform PCA with the packages *xcms* (3.2.0)[99], *FactoMineR* (2.3)[105], and *MetabolAnalyze* (1.3.1).[103, 104] Settings for processing in R were as follows: method, centWave; prefilter, (5, 5000); ppm, 5; snthresh, 10; peakwidth, (5, 20); noise, 500000; bw, 5; minfrac, 0.001; mzwid, 0.015. Zero values were imputed with two-thirds of the lowest value measured for each metabolite. Peak area values were log-10 transformed and pareto scaled. PCA was performed with QC samples to assess instrument drift over the course of analysis. The first and second principal components were plotted against one another, because they account for the most variability within a dataset.

3.3.3 GNPS parameters and processing

The mzML files generated for PCA were uploaded to GNPS with FileZilla (3.9.0.5) and analyzed with the following settings: precursor ion mass tolerance, 0.02 Da; fragment ion mass tolerance, 0.02 Da; min pairs cosine, 0.75; network topK, 10; maximum connected component size, 100; minimum matched fragment ions, 5; minimum cluster size, 2; MSCluster, on. The network output was downloaded as a GRAPHML file to import into Cytoscape (3.6.1) for visualization. To simplify network analysis, all unconnected nodes were removed, along with nodes attributed to media files, clusters formed solely from seed spectra, and background ions.

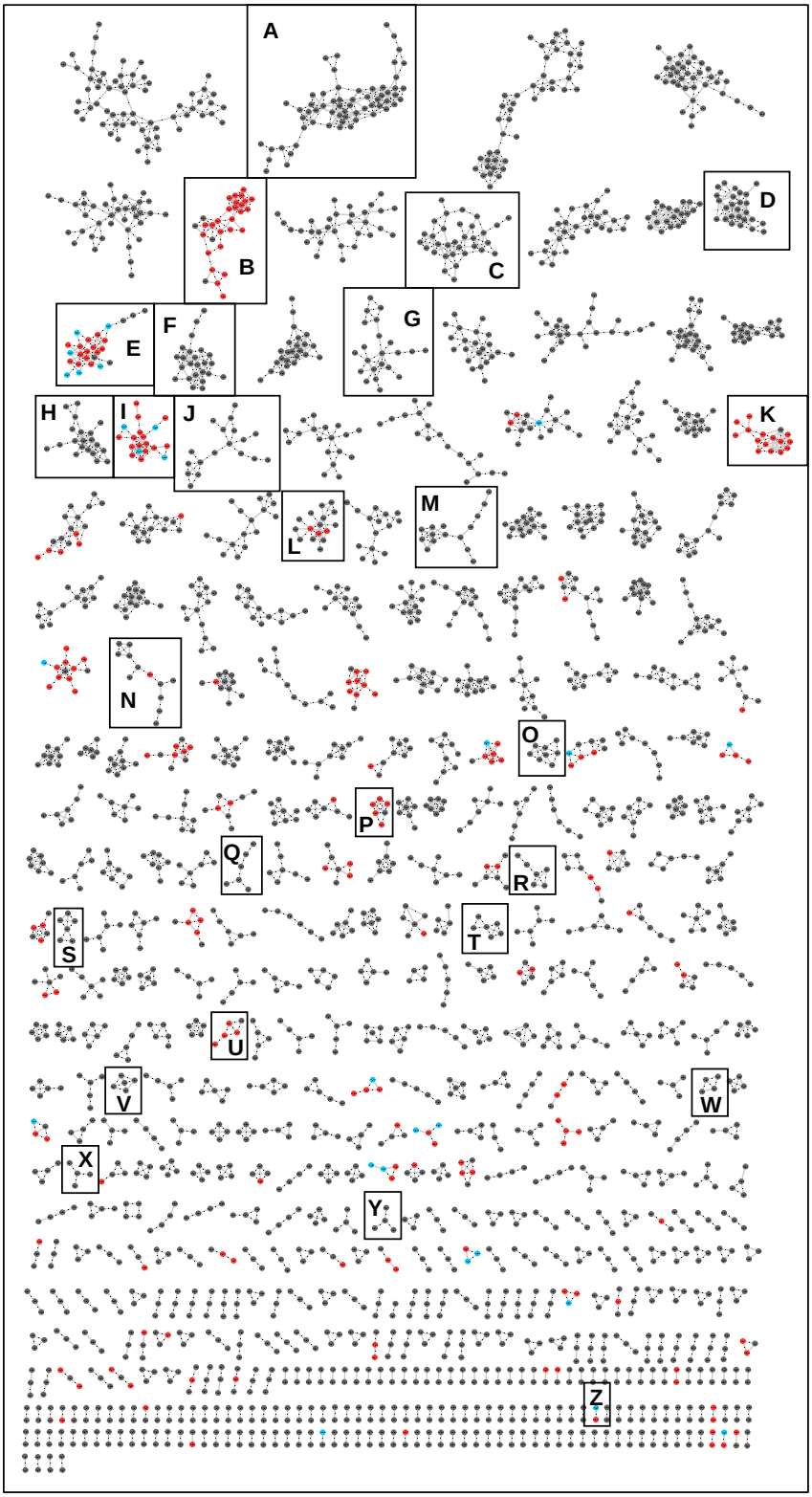
Files were also assessed with the Library Search function of GNPS to dereplicate known compounds. The following parameters were used: precursor ion mass tolerance, 0.02 Da; fragment ion mass tolerance, 0.02 Da; min matched fragment ions, 5; cut-off score, 0.75. Compounds were additionally dereplicated by comparing to an in-house database of MS/MS spectra.

3.4 Results and Discussion

Samples were analyzed solely in positive ionization mode because in preliminary tests, more features were successfully ionized and fragmented in positive mode versus negative mode. PCA was performed primarily to assess instrumental drift over the course of analysis. QC samples all grouped tightly together, indicating little to no instrumental drift (see Appendix C, Figure C1). Upon initial creation of the network, there were 12,037 nodes with 16,223 connections. After eliminating unpaired nodes, features from media, and features with poor chromatography, there remained 2804 nodes with 4119 connections. Along with generation of the molecular network, files were assessed with the GNPS library search tool, to dereplicate compounds within the raw

files. The full molecular network is pictured below, in Figure 22, with clusters containing dereplicated compounds labeled.

The GNPS library contains over 77,000 spectra to compare with, however, this is still only a fraction of the natural products that have been reported. Nineteen secondary metabolites were successfully dereplicated by the library search function. To improve the number of dereplicated compounds, nodes were also compared with an in-house library of MS/MS spectra, on an individual basis. In total, 60 compounds were dereplicated within the network and are listed in Table 5.



- **Seed spectra**
- **Seed and endophyte**
- **Endophyte only**

Compounds:

A. Hormonemate
B. Cytochalasins
C. Enniatins
D. Statins
E. Ellisiiamides
F. Pyridone toxins
G. Sclerodins
H. [3-methyl-1-[3-methyl-1-[3-methyl-1-[3-methyl-1-oxo-1-(2,3,4,5-tetrahydroxypentoxy)pentan-2-yl]oxy-1-oxopentan-2-yl]oxy-1-oxopentan-2-yl]oxy-1-oxopentan-2-yl] 2-acetyloxy-3-methylpentanoate
I. Coleophomone intermediate
J. (E)-4-(6,8-dihydroxy-3-methyl-1-oxoisochromen-7-yl)-2-methoxybut-3-enoic acid
K. Cytochalasins
L. Griseofulvin
M. Pyrones
N. Senkyunolide H
O. Trichodin A
P. Hirsutatin A
Q. Trienylfuranol A
R. LL P880 beta
S. Lachnellulones
T. Lovastatin hydroxy acid
U. Griseofernaneosides
V. Akanthomycin analogue
W. Ergochromes
X. Lachnochromonin A
Y. Echinulins
Z. Fulvic acid derivative

Figure 22. Continued

Figure 22. (Above) Molecular network of LC-HRMS features from Canadian fungal endophytes, generated with GNPS. Nodes represent LC-MS features and lines connecting them indicate features that share a cosine similarity score higher than 0.75. Clusters that contain dereplicated structural groups are labeled.

With a large network comprising many unique fungal species and LC-MS features, an additional approach to identifying new compounds was necessary, other than attempting to dereplicate every single node. Attention was given to clusters containing dereplicated nodes, because all seed compounds that were included and many compounds found by the library search are bioactive to some degree. Therefore, unknown compounds within clusters that contain a known bioactive node are likely also bioactive and worthwhile to pursue. Clusters containing known compounds were examined closely to determine the identities of as many nodes as possible.

Dereplicated compounds found within the network encompass a broad range of compound types. These can be observed below in Table 5, organized by cluster, and their structures are shown in Figure 23. Select clusters that were used to identify new and novel compounds are examined in detail in sections 3.4.1-3.4.5 below.

Table 5. Natural products dereplicated from molecular network of 302 Canadian fungal endophytes. Formulas listed do not include adducts.

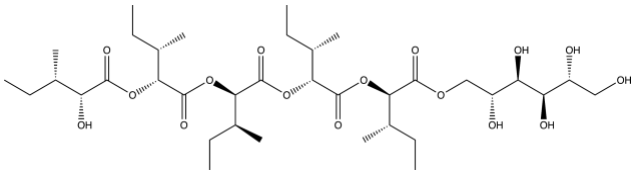
Cluster	Name	Measured *m/z*	Formula	RT (min)	Mass error (Δppm)	Species
A	Hormonemate F	626.3747 $[M+NH_4]^+$	$C_{29}H_{52}O_{13}$	4.38	0.116	*Ramularia and Rhizosphaera spp.*
	Hormonemate A/E	740.4426 $[M+NH_4]^+$	$C_{35}H_{62}O_{15}$	4.77	-0.171	*Ramularia and Rhizosphaera spp.*
	Hormonemate[a]	770.4531 $[M+NH_4]^+$	$C_{36}H_{64}O_{16}$	4.67	-0.170	*Ramularia and Rhizosphaera spp.*
B	Phenochalasin C	450.2639 $[M+H]^+$	$C_{28}H_{35}NO_4$	4.03	-0.145	*Xylaria ellisii*
	Cytochalasin Z26/Z22/Z23	480.2382 $[M+H]^+$	$C_{28}H_{33}NO_6$	3.72	0.199	*Xylaria ellisii*
	Zygosporin E	492.2744 $[M+H]^+$	$C_{30}H_{37}NO_5$	4.14	0.244	*Xylaria ellisii*
	Epoxycytochalasin C/D	524.2645 $[M+H]^+$	$C_{30}H_{37}NO_7$	3.81	0.479	*Xylaria ellisii*
	Epoxycytochalasin N/R	540.2593 $[M+H]^+$	$C_{30}H_{37}NO_8$	3.72	0.160	*Xylaria ellisii*
	Cytochalasin P1	542.2749 $[M+H]^+$	$C_{30}H_{39}NO_8$	3.54	-0.007	*Xylaria ellisii*
C	Enniatin J1	629.4120 $[M+NH_4]^+$	$C_{31}H_{53}N_3O_9$	4.78	0.070	*Fusarium tricinctum*
	Enniatin B2/J2/K	643.4276 $[M+NH_4]^+$	$C_{32}H_{55}N_3O_9$	4.91	0.224	*Fusarium tricinctum*
	Enniatin B	657.4443 $[M+NH_4]^+$	$C_{33}H_{57}N_3O_9$	5.00	1.010	*Fusarium tricinctum*
	Enniatin P1	659.4223 $[M+NH_4]^+$	$C_{32}H_{55}N_3O_{10}$	4.55	-0.410	*Fusarium tricinctum*
	Enniatin P2	673.4382 $[M+NH_4]^+$	$C_{33}H_{57}N_3O_{10}$	4.73	-0.075	*Fusarium tricinctum*
	Enniatin A1/E1/E2/G/O2	685.4746 $[M+NH_4]^+$	$C_{35}H_{61}N_3O_9$	5.23	0.269	*Fusarium tricinctum*
	Enniatin A/A2/C/F/MK1688	699.4902 $[M+NH_4]^+$	$C_{36}H_{63}N_3O_9$	5.38	-0.209	*Fusarium tricinctum*
	Enniatin M1/M2	701.4698 $[M+NH_4]^+$	$C_{35}H_{61}N_3O_{10}$	4.94	0.470	*Fusarium tricinctum*
D	Lovastatin[a]	405.2631 $[M+H]^+$	$C_{24}H_{36}O_5$	4.41	-1.211	*Seimatosporium lichenicola*
	Simvastatin[a]	419.2793 $[M+H]^+$	$C_{25}H_{38}O_5$	4.79	-0.073	*Seimatosporium lichenicola*
	Lovastatin analogue[a]	482.3110 $[M+NH_4]^+$	$C_{26}H_{40}O_7$	4.09	-0.392	*Seimatosporium lichenicola*

E	Ellisiiamide D	522.3649 $[M+H]^+$	$C_{27}H_{47}N_5O_5$	4.29	-0.012	*Xylaria ellisii*
	Cyclic pentapeptide 2	536.3804 $[M+H]^+$	$C_{28}H_{49}N_5O_5$	4.16	-0.515	*Xylaria ellisii*
	Xylarotide A	550.3968 $[M+H]^+$	$C_{29}H_{51}N_5O_5$	4.45	0.225	*Xylaria ellisii*
	Ellisiiamide A	556.3494 $[M+H]^+$	$C_{30}H_{45}N_5O_5$	4.29	0.043	*Xylaria ellisii*
	Ellisiiamide B	570.3649 $[M+H]^+$	$C_{31}H_{47}N_5O_5$	4.32	-0.116	*Xylaria ellisii*
	Cyclic pentapeptide 1	584.3807 $[M+H]^+$	$C_{32}H_{49}N_5O_5$	4.50	0.161	*Xylaria ellisii*
	Ellisiiamide C	598.3961 $[M+H]^+$	$C_{33}H_{51}N_5O_5$	4.61	-0.294	*Xylaria ellisii*
	Ellisiiamide G	600.3754 $[M+H]^+$	$C_{32}H_{49}N_5O_6$	4.11	-0.218	*Xylaria ellisii*
F	Demethylsambutoxin	440.2793 $[M+H]^+$	$C_{27}H_{37}NO_4$	4.82	-0.534	*Fusarium tricinctum*
	Sambutoxin	454.2952 $[M+H]^+$	$C_{28}H_{39}NO_4$	5.07	-0.033	*Fusarium tricinctum*
	Anhydrooxysporidinone	472.3059 $[M+H]^+$	$C_{28}H_{41}NO_5$	4.73	0.403	*Fusarium tricinctum*
	Oxysporidinone[a]	490.3162 $[M+H]^+$	$C_{28}H_{43}NO_6$	4.69	-0.132	*Fusarium tricinctum*
G	Trypethelone	273.1120 $[M+H]^+$	$C_{16}H_{16}O_4$	4.02	-0.496	*Godronia cassandrae*
	Sclerodin	329.1019 $[M+H]^+$	$C_{18}H_{16}O_6$	4.44	-0.288	*Godronia cassandrae*
	Sclerodinol	345.0969 $[M+H]^+$	$C_{18}H_{16}O_7$	4.28	0.292	*Godronia cassandrae*
H	Coleophomone intermediate	469.1857 $[M+H]^+$	$C_{26}H_{28}O_8$	3.65	0.076	*Xylaria cubensis*
I	[3-methyl-1-[3-methyl-1-[3-methyl-1-[3-methyl-1-oxo-1-(2,3,4,5-tetrahydroxypentoxy)pentan-2-yl]oxy-1-oxopentan-2-yl]oxy-1-oxopentan-2-yl]oxy-1-oxopentan-2-yl] 2-acetyloxy-3-methylpentanoate[a]	782.4527 $[M+NH_4]^+$	$C_{37}H_{64}O_{16}$	5.04	-0.641	*Rhizosphaera spp.*

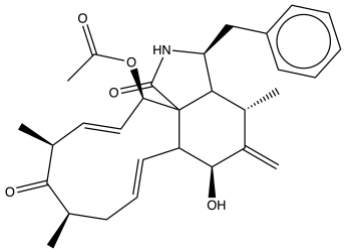
J	(E)-4-(6,8-dihydroxy-3-methyl-1-oxoisochromen-7-yl)-2-methoxybut-3-enoic acid[a]	307.0812 $[M+H]^+$	$C_{15}H_{14}O_7$	3.76	-0.258	*Ramularia, Epicoccum, and Mollisia* spp.
K	Cytochalasin D	508.2696 $[M+H]^+$	$C_{30}H_{37}NO_6$	3.93	-0.427	*Xylaria ellisii*
L	Griseophenone C	305.1019 $[M+H]^+$	$C_{16}H_{16}O_6$	3.21	-0.212	*Xylaria ellisii*
	Dechlorogriseofulvin[a]	319.1179 $[M+H]^+$	$C_{17}H_{18}O_6$	3.79	1.144	*Xylaria ellisii*
	Griseofulvin[a]	353.0786 $[M+H]^+$	$C_{17}H_{17}ClO_6$	3.97	0.078	*Xylaria ellisii*
M	(E)-5-(4-methoxy-5-methyl-6-oxopyran-2-yl)-3-methylhex-4-enoic acid[a] (similar to marinopyrones)	267.1226 $[M+H]^+$	$C_{14}H_{18}O_5$	3.65	-0.300	Xylariales order
N	Senkyunolide H[a]	225.1122 $[M+H]^+$	$C_{12}H_{16}O_4$	3.56	0.153	*Nigrospora sphaerica* and *Cytospora ribis*
O	Trichodin A analogue[a]	354.2063 $[M+H]^+$	$C_{22}H_{27}NO_3$	3.73	-0.452	*Cryptosporella femoralis*
P	LL P880 beta (fungal lactone)[a]	231.1228 $[M+H]^+$	$C_{11}H_{18}O_5$	3.15	0.389	*Xylaria ellisii*
Q	Trienylfuranone A/B	179.1067 $[M+H]^+$	$C_{11}H_{14}O_2$	2.79	-0.091	*Hypoxylon submonticulosum*
	Trienylfuranol A	181.1223 $[M+H]^+$	$C_{11}H_{16}O_2$	3.72	0.131	*Hypoxylon submonticulosum*
R	Hirsutatin A	677.3756 $[M+H]^+$	$C_{34}H_{52}N_4O_{10}$	4.38	-0.104	*Xylaria ellisii*
S	Deoxylachnellulone	309.2059 $[M+H]^+$	$C_{18}H_{28}O_4$	4.59	-0.342	*Lachnellula calyciformis*
	Lachnellulone	325.2009 $[M+H]^+$	$C_{18}H_{28}O_5$	4.37	-0.063	*Lachnellula calyciformis*
T	Lovastatin hydroxy acid[a]	445.2554 $[M+Na]^+$	$C_{24}H_{38}O_6$	4.41	-1.415	*Seimatosporium lichenicola*
U	Griseofernaneoside B	647.4149 $[M+H]^+$	$C_{37}H_{58}O_9$	3.98	-0.587	*Xylaria ellisii*
V	Akanthomycin analogue[a]	372.2169 $[M+H]^+$	$C_{22}H_{29}NO_4$	4.22	-0.040	*Leptodontidium sp.*

W	Ergochrome EE/secalonic acid[a]	639.1706 $[M+H]^+$	$C_{32}H_{30}O_{14}$	4.36	-0.379	*Coniothyrium ferrarisianum*
	Ergochrome AD/BD	657.1811 $[M+H]^+$	$C_{32}H_{32}O_{15}$	3.95	-0.467	*Coniothyrium ferrarisianum*
	Ergochrome DD	675.1917 $[M+H]^+$	$C_{32}H_{34}O_{16}$	3.62	-0.283	*Coniothyrium ferrarisianum*
X	Preechinulin[a]	326.1863 $[M+H]^+$	$C_{19}H_{23}N_3O_2$	3.57	-0.164	*Aspergillus amstelodami*
	Tardioxopiperizine A/B	394.2489 $[M+H]^+$	$C_{24}H_{31}N_3O_2$	4.22	0.067	*Aspergillus amstelodami*
	Echinulin[a]	462.3115 $[M+H]^+$	$C_{29}H_{39}N_3O_2$	4.78	-0.030	*Aspergillus amstelodami*
	Hydroxyechinulin	478.3065 $[M+H]^+$	$C_{29}H_{39}N_3O_3$	4.83	-0.499	*Aspergillus amstelodami*
Y	Lachnochromonin A	263.1279 $[M+H]^+$	$C_{15}H_{18}O_4$	3.45	0.777	*Lachnum* sp.
Z	7-hydroxy-3-(hydroxymethyl)-2-(2-hydroxypropyl)-6-methoxy-4H-chromen-4-one (fulvic acid derivative)	281.1021 $[M+H]^+$	$C_{14}H_{16}O_6$	3.01	0.090	*Sphaerulina rhabdoclinis*

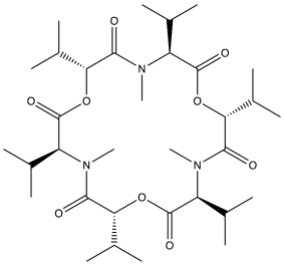
[a] Dereplicated with GNPS Library search function



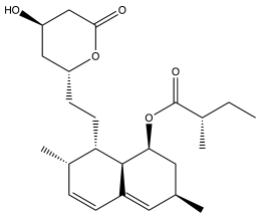
Hormonemate
Cluster A



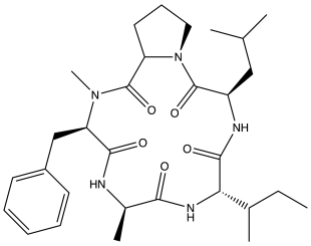
Zygosporin E
Cluster B



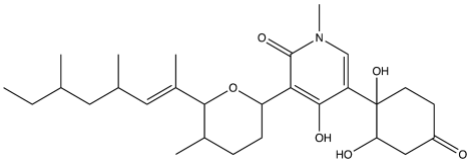
Enniatin B
Cluster C



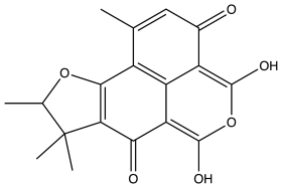
Lovastatin
Cluster D



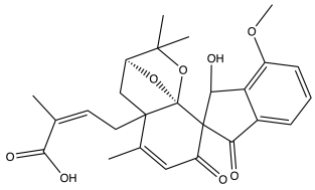
Ellisiiamide A
Cluster E



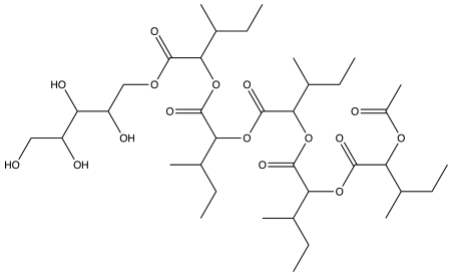
Oxysporidinone
Cluster F



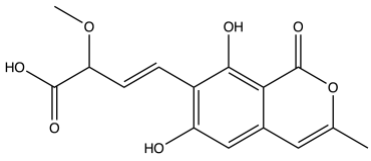
Sclerodin
Cluster G



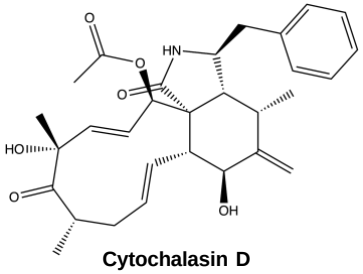
Coleophomone intermediate
Cluster H



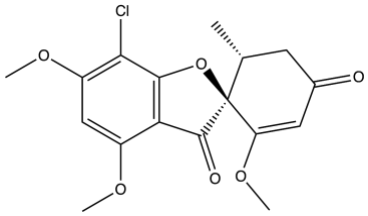
[3-methyl-1-[3-methyl-1-[3-methyl-1-[3-methyl-1-oxo-1-(2,3,4,5-tetrahydroxypentoxy)pentan-2-yl]oxy-1-oxopentan-2-yl]oxy-1-oxopentan-2-yl]oxy-1-oxopentan-2-yl] 2-acetyloxy-3-methylpentanoate
Cluster I



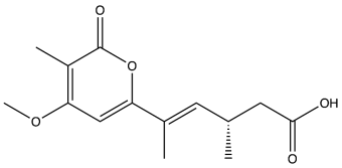
(E)-4-(6,8-dihydroxy-3-methyl-1-oxoisochromen-7-yl)-2-methoxybut-3-enoic acid
Cluster J



Cytochalasin D
Cluster K



Griseofulvin
Cluster L



(E)-5-(4-methoxy-5-methyl-6-oxopyran-2-yl)-3-methylhex-4-enoic acid
Cluster M

Figure 23. Continued

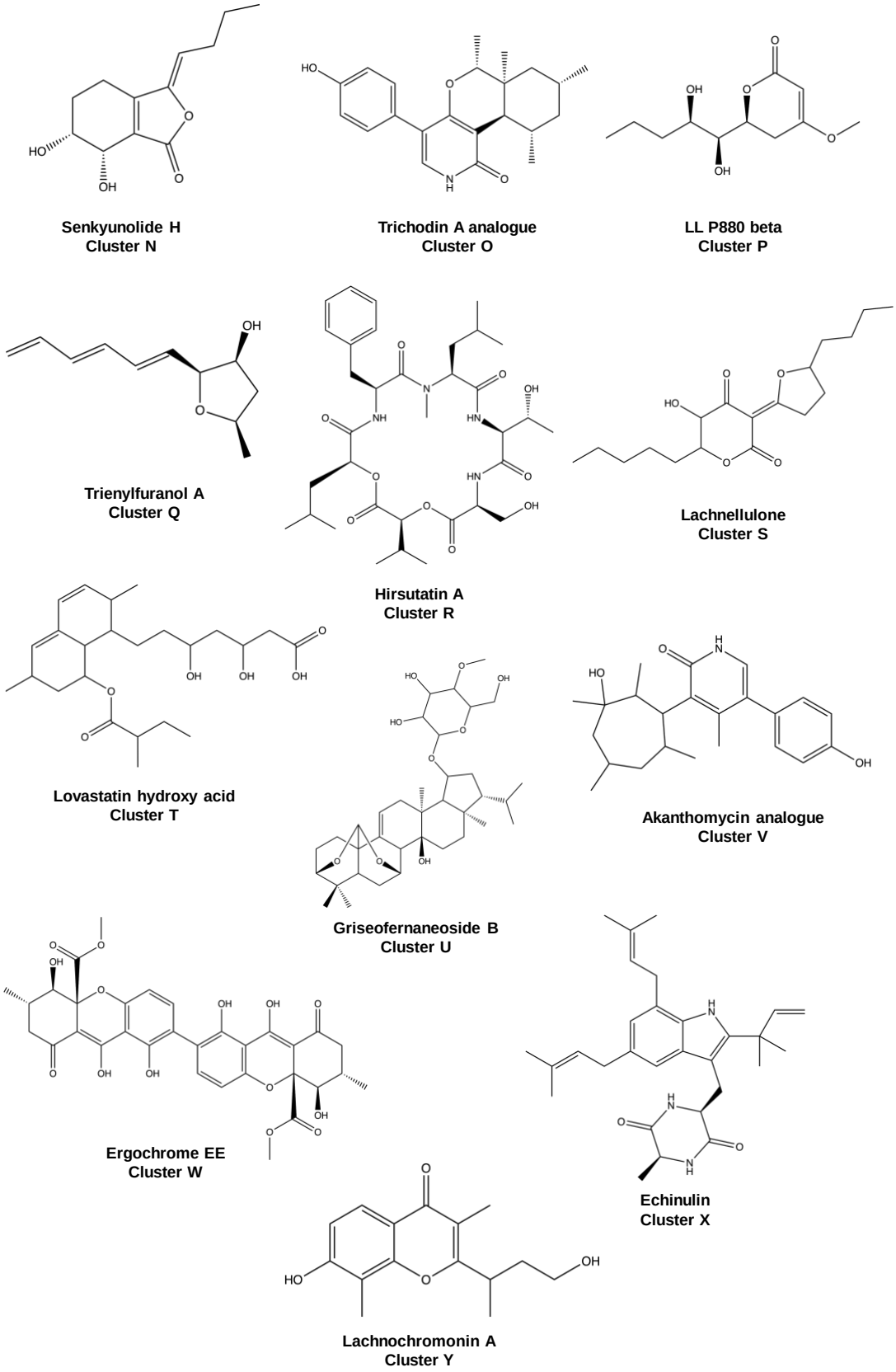


Figure 23. Continued

Figure 23. (Above) Representative structures of compounds dereplicated from clusters within the spectral molecular network shown in Figure 22.

3.4.1 Griseofulvin Cluster

Within cluster L, both griseofulvin and dechlorogriseofulvin were dereplicated by the GNPS library search function. Both are highlighted below in Figure 24. The *M + 2* node is present for griseofulvin and represents its chlorine isotope, providing further confidence in this assignment. Two other nodes connected to griseofulvin have *m/z* values of 305.1019 and 335.1125, representing chemical formulas of $C_{16}H_{16}O_6$ and $C_{17}H_{18}O_7$, respectively. The compound with the formula of $C_{16}H_{16}O_6$ is tentatively identified as griseophenone C, a relative of griseofulvin. It has a prominent product ion of 165.0546, which is common to griseofulvin and griseophenone B.[131] It is connected with griseofulvin and dechlorogriseofulvin by cosine scores of 0.93 and 0.94, respectively. The compound with the formula $C_{17}H_{18}O_7$ (**1**) shares the same product ion of *m/z* 165.0546 (see Figure 25), but its formula was not comparable to any previously published griseofulvin-related compounds. It has cosine scores of 0.94 and 0.93 connecting it with griseofulvin and dechlorogriseofulvin, indicating high degrees of spectral similarity among these compounds. All nodes within bolded circles in Figure 24 were detected in several extracts of *Xylaria ellisii*.

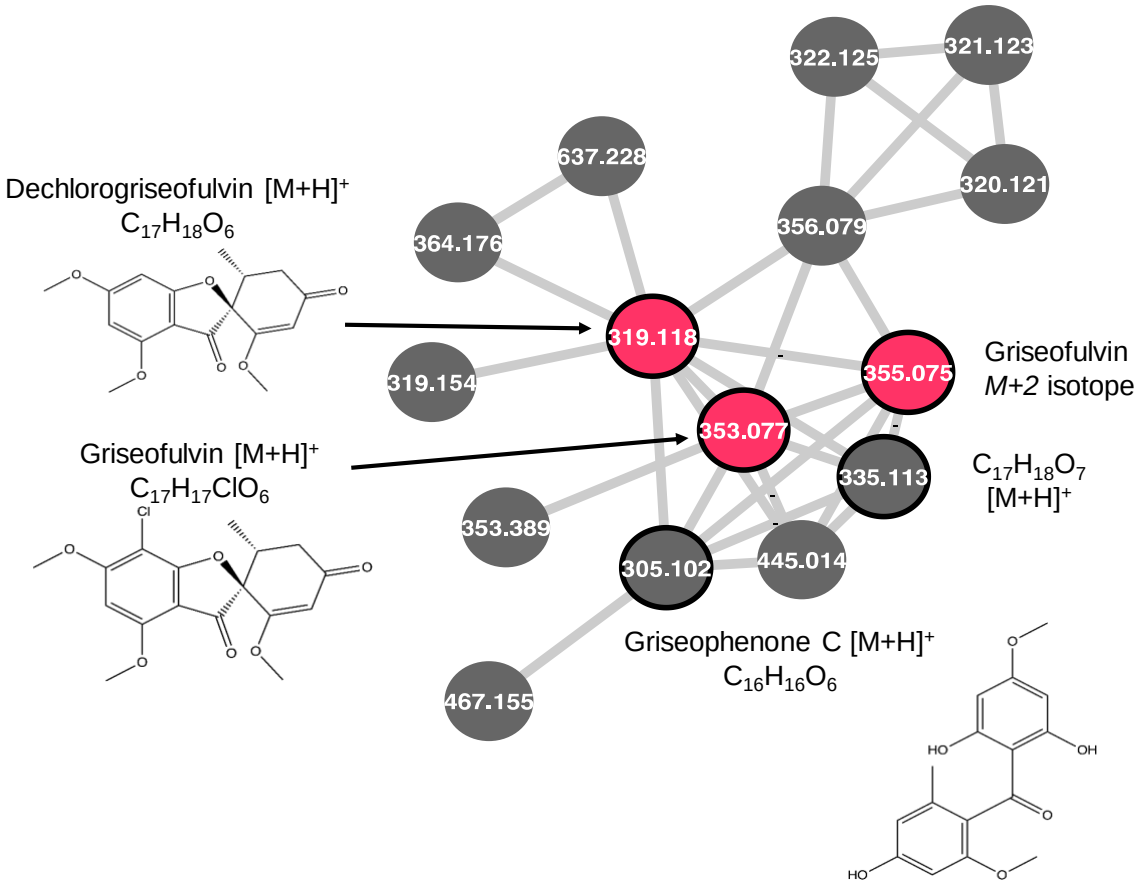


Figure 24. Close-up of cluster L from molecular network of 302 fungal endophytes, containing griseofulvin and related compounds.

Griseofulvin is a potent antifungal compound that is used as a broad-spectrum fungicide in treating human fungal infections, and has shown promise in treating other conditions such as gout and ischemic heart disease.[132, 133] It was first isolated from *Penicillium griseofulvum*, and is a known metabolite of *Xylaria ellisii*.[133, 134] Its dechlorinated analogue also shows antifungal activity, albeit to a weaker degree.[135, 136] Griseophenone C is a known precursor in griseofulvin's biosynthetic pathway and is also a potent antimicrobial agent, particularly against methicillin-resistant *Staphylococcus aureus* (MRSA) and *Escherichia coli*.[137]

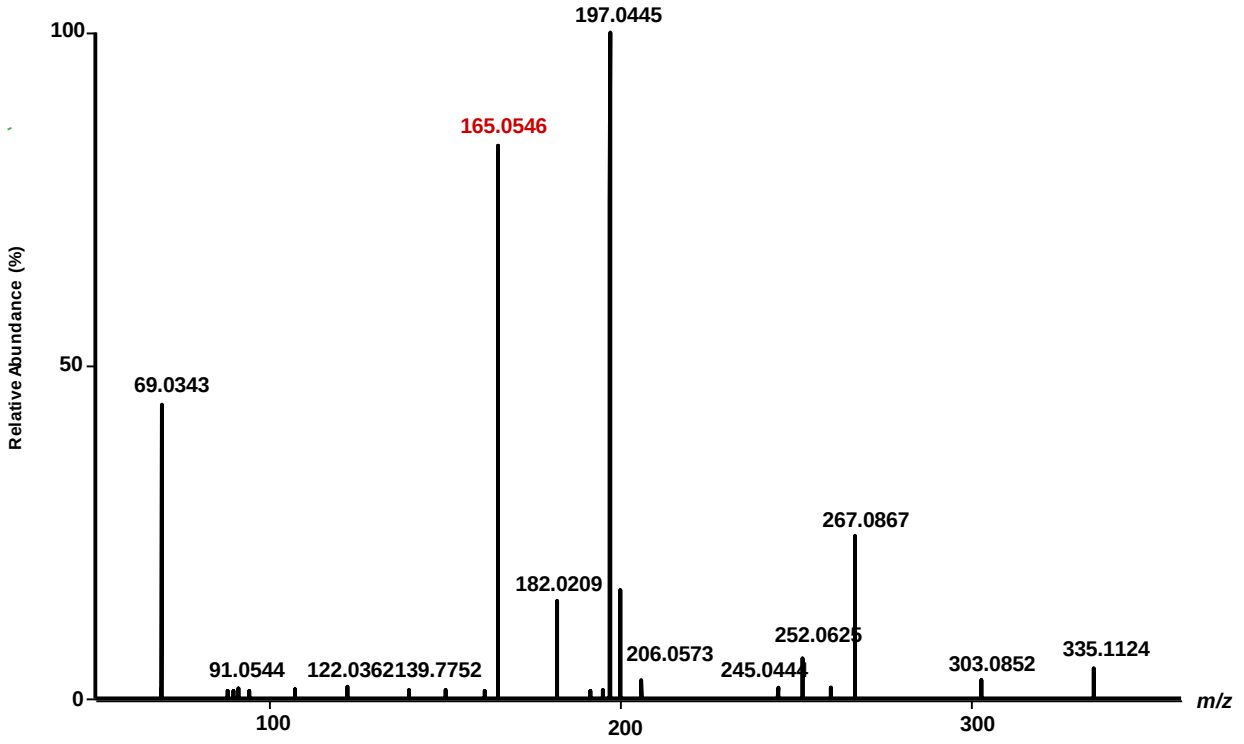


Figure 25. Tandem mass spectrum of an unknown griseofulvin-related compound with the chemical formula $C_{17}H_{18}O_7$. A fragment characteristic to griseofulvin is highlighted in red.

One may expect to see other griseophenones in cluster L, because they are also part of the biosynthetic pathway to form griseofulvin and dechlorogriseofulvin. However, no other formulas matching those of reported griseophenones were present. There were unpaired nodes with *m/z* values matching those of griseophenone B and D present in the network before unpaired nodes were removed but manual inspection of their tandem mass spectra revealed poor fragmentation at this collision energy, giving them too few fragments to compare with other spectra. Because they were unpaired, they were removed from the network.

The antimicrobial activities demonstrated by griseofulvin and its relatives make the unidentified compound an excellent target for isolation and characterization, because the unknown compound may also share some bioactivity.

3.4.2 Griseofernaneoside Cluster

Cluster U contains griseofernaneoside B, which was dereplicated by comparison to a published tandem mass spectrum.[134] Griseofernaneosides A and B were first reported from *Xylaria ellisii* and are unique fernane-type triterpenes.[134] This compound class is mostly isolated from ferns and other plants, and is very uncommon in microbes.[134] Griseofernaneoside B is present in cluster M with both a protonated and an ammoniated ion (see Figure 26, below). A feature with a *m/z* matching that of griseofernaneoside A was present in raw files from *Xylaria ellisii* strains, however its abundance was below the AGC target threshold and it was not fragmented by MS/MS.

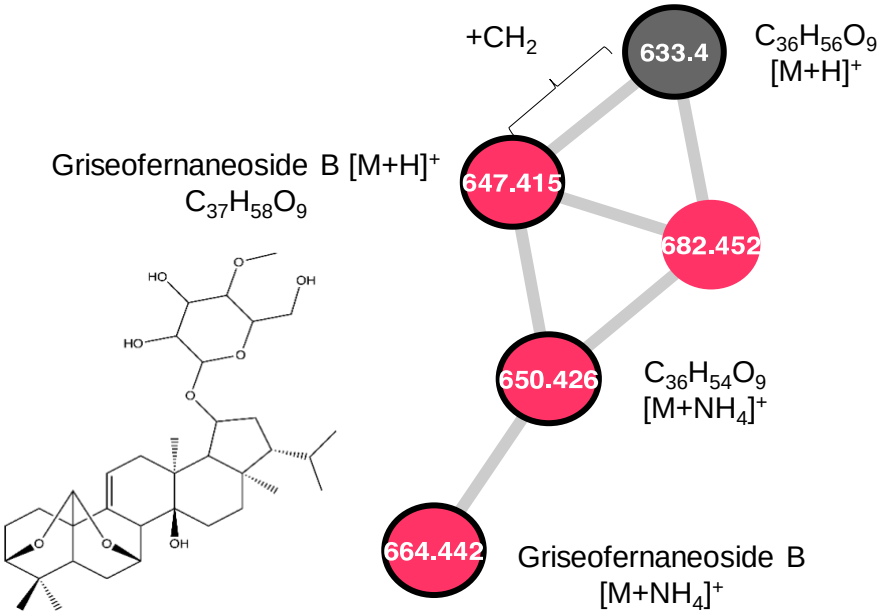


Figure 26. Close-up of cluster U, which contains griseofernaneoside B and related compounds.

There are also two unknown compounds with *m/z* values of 650.4262 $[M+H]^+$ and 633.3995 $[M+NH_4]^+$ that had formulas of $C_{36}H_{54}O_9$ (**2**) and $C_{36}H_{56}O_9$ (**3**), respectively. Both have well-fragmented spectra that are highly similar to that of griseofernaneoside B (see Figure 27, below). A search of these formulas through SciFinder and Antibase did not yield any other reported

compounds from the family of fernane terpenes. The $[M+H]_+$ node of griseofernaneoside B has a high cosine similarity score of 0.85 with compound **3**, and a score of 0.75 with compound **2**. The unknown compounds share a similar number, type, and intensity of fragments with griseofernaneoside B, which indicates a high degree of structural similarity. All bolded nodes within cluster U were solely made by *Xylaria ellisii*.

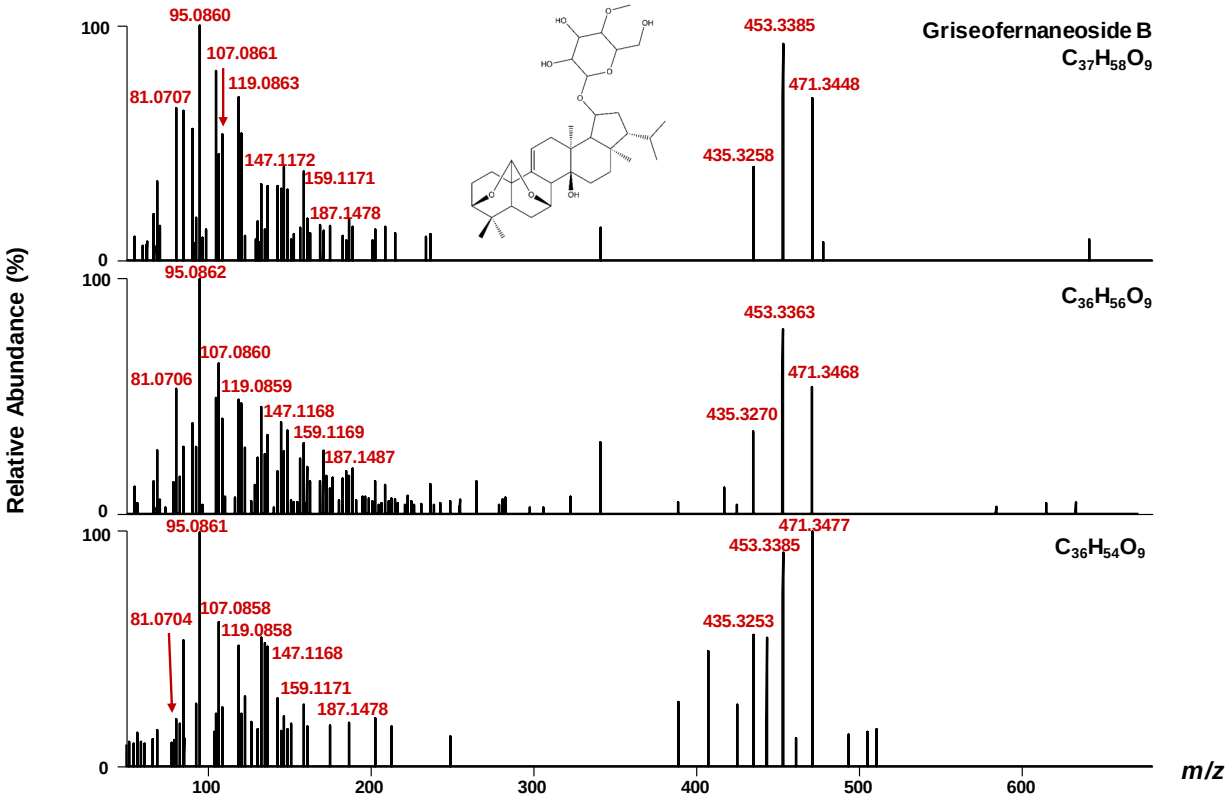


Figure 27. Tandem mass spectrum of griseofernaneoside B and two related compounds with the chemical formulas $C_{36}H_{56}O_9$ and $C_{36}H_{54}O_9$. Fragments that are common to all three compounds are labeled in red.

The bioactivities of griseofernaneosides A and B have not been assessed, but a compound with a similar core structure isolated from ferns has shown mild activity *in vitro* against two human cancer cell lines.[138] Compounds **2** and **3** are likely two new griseofernaneoside-related compounds that are good targets for isolating and characterizing.

3.4.3 Hirsutatin A Cluster

Hirsutatin A is a cyclohexadepsipeptide that was originally discovered from an insect pathogenic fungus, but was later isolated from *Xylaria ellisii*.[134, 139] Within the molecular network, it was dereplicated from the network by comparison with the included seed spectrum file. Hirsutatin A is present in crude extracts of *Xylaria ellisii* as well from one extract each of *Xylaria cubensis* and *Godronia cassandrae*. It has both protonated $[M+H]^+$ and ammoniated $[M+NH_4]^+$ ions in cluster R. It is connected to three nodes that could not be identified, having chemical formulas of $C_{34}H_{52}N_4O_9$ (**4**), $C_{35}H_{54}N_4O_{10}$ (**5**), and $C_{36}H_{59}N_5O_{10}$ (**6**). These shared cosine similarity scores with Hirsutatin A of 0.84, 0.88, and 0.95, respectively.

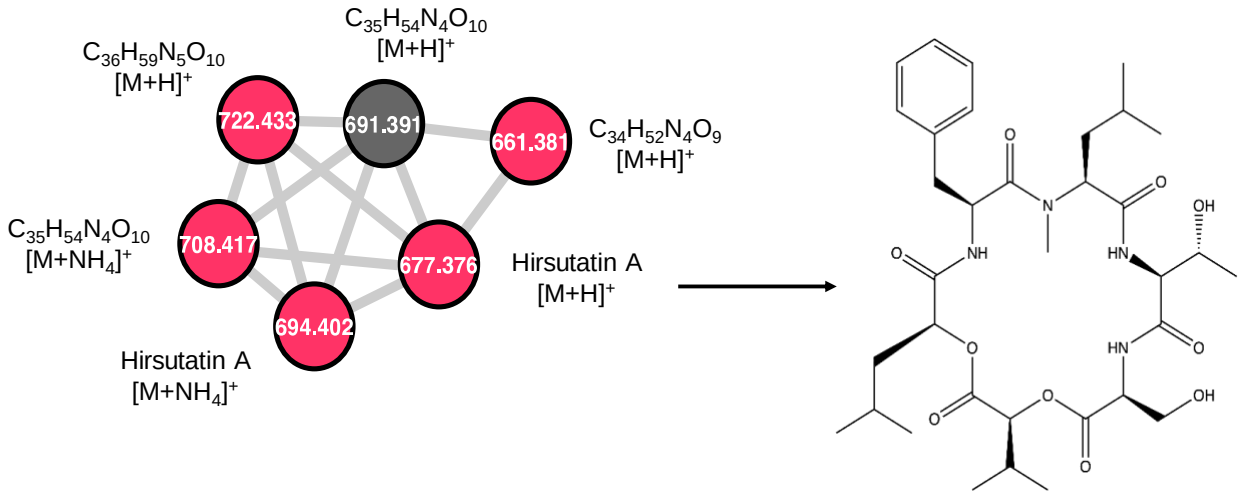


Figure 28. Close-up of cluster R, containing hirsutatin A and three unidentified related compounds.

Hirsutatin A has mild activity against *Mycobacterium tuberculosis*, which is the causative agent of tuberculosis in humans.[139] Its close relative, hirsutatin B ($C_{35}H_{54}N_4O_{11}$) was not present in the molecular network or in raw files. Hirsutatin B has strong activity against a multi-drug resistant strain of *Plasmodium falciparum* and mild activity against *M. tuberculosis*.[139] *P.*

falciparum is the parasite that causes malaria and is responsible for hundreds of thousands of deaths annually. Other hirsutatin-related compounds also have the potential for possessing bioactivity against tuberculosis and malaria pathogens, making the compounds within cluster R worthy of examining further.

3.4.4 Oxysporidinone Cluster

Oxysporidinone was dereplicated via the GNPS library search. Oxysporidinone is an antifungal compound that was originally reported from *Fusarium oxysporum*, and here was detected from *Fusarium tricinctum*.[140] Breinholt et al. reported that oxysporidinone has strong activity against several plant pathogenic fungi, including *Aspergillus niger*, *Botrytis cinerea*, *Alternaria alternata*, and *Venturia inequalis*.[140] A search of chemical formulas within cluster F yielded several compounds with similar chemical formulas, including a compound matching the formula of sambutoxin. Sambutoxin is a mycotoxin with documented toxicity in rats and is also a known product of *Fusarium* spp.[141, 142] There are two additional nodes in the cluster with *m/z* values of 440.2793 and 472.3059, corresponding to chemical formulas of $C_{27}H_{37}NO_4$ and $C_{28}H_{41}NO_5$. These match the formulas of demethylsambutoxin and anhydrooxysporidinone, respectively, and are tentatively identified as such. Both demethylsambutoxin and anhydrooxysporidinone are known products of *Fusarium* spp., but neither have reported bioactivity against the small range of organisms they have been tested against.[143] Other related compounds made by *Fusarium* spp. such as the antibacterial fusapyridons were not present.[144]

Cluster F also contains two nodes whose chemical formulas did not match any published compounds that are structurally related to oxysporidinone or sambutoxin. They had experimental *m/z* values of 438.3001 and 504.3321 which indicate formulas of $C_{28}H_{39}NO_3$ (**7**) and $C_{29}H_{45}NO_6$ (**8**), respectively. Their tandem mass spectra are included in Appendix C.

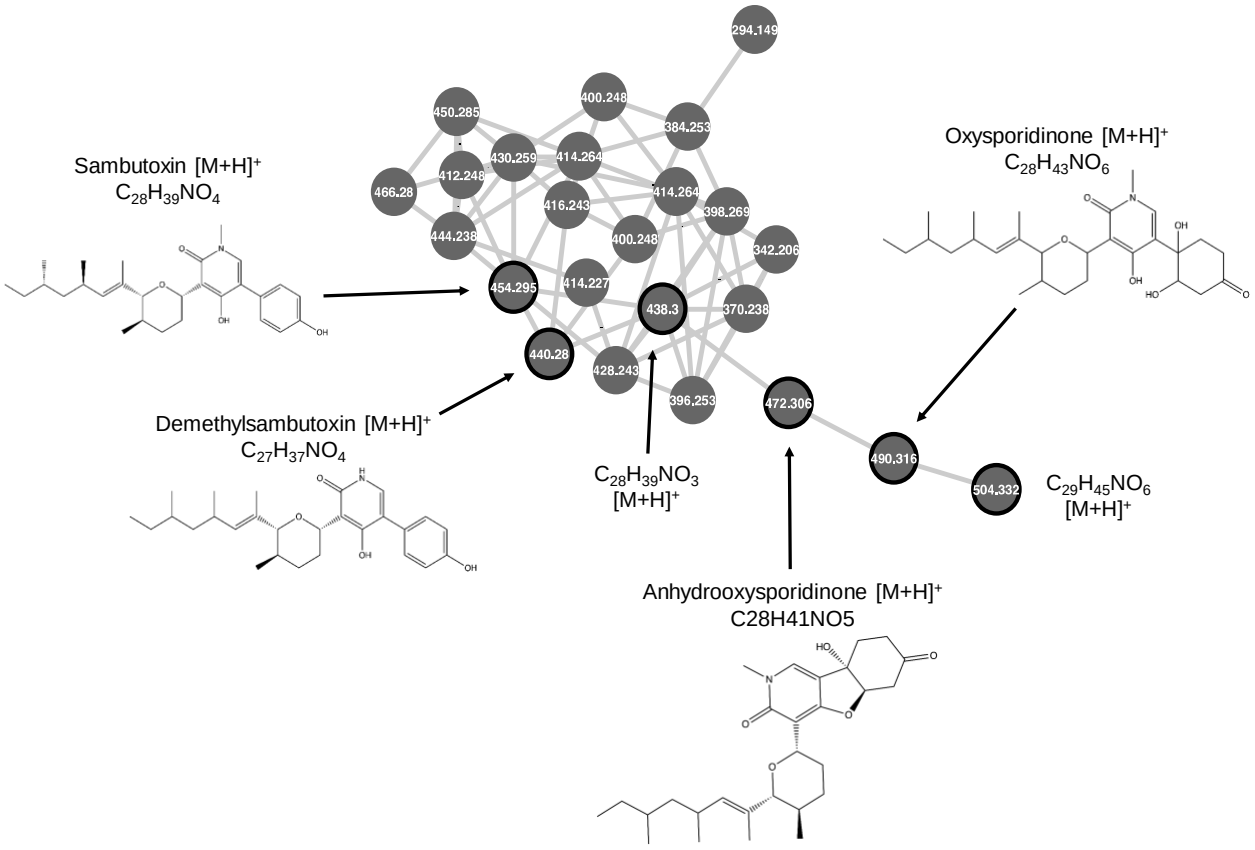


Figure 29. Close-up of cluster F, which contains several known mycotoxins as well as two unknown entities.

The range of bioactivities exhibited by these pyridone alkaloids make the unknown compounds interesting targets for closer examination.

3.4.5. Additional Dereplicated Compounds

The GNPS library search successfully dereplicated a handful of secondary metabolites, many from previously reported sources. For example, echinulin was identified here solely in extracts of *Aspergillus amstelodami*. Echinulin was first reported as a product from *Aspergillus glaucus* in 1948 and has since been reported from several other species of *Aspergillus*, including *Aspergillus amstelodami*.[145] Many known products of *Xylaria ellisii* were also identified, such as the ellisiiamides, hirsutatin A, and several cytochalasins. Most seed spectra compounds were

present within the network, except for coriloxin and ascochitine, abscisic acid, and the fulvic acid analogue. These were present in raw files of species that are known to make them, but they did not appear in the network because there were no related compounds they could connect to, or their fragmentation was too poor.

In addition to known fungal sources, there are also some dereplicated compounds that appear to be made by novel sources. Along with finding new compounds, discovering novel sources of known natural products is also worthwhile. Fungi differ widely in the abundance of compounds they can produce, how easy they are to culture in a lab, and in their relationships to hosts. Finding new sources of known natural products can find more efficient ways of producing those products or may illuminate answers of chemical ecology.

Lovastatin is a cholesterol-reducing drug that was originally isolated from the citrus pathogen *Penicillium citrinum*.[146] Here it was present in two distinct extracts of *Seimatosporium lichenicola*, which resides in berry plants such as currants.[147] Secalonic acid D is commonly produced by *Aspergillus* and *Penicillium* spp. as a foil spoilage toxin but here it is present in an extract of *Coniothyrium ferrarisianum*.[148, 149] Other species of *Coniothyrium* are plant and human pathogens, as well as parasites of fungal pathogens[150]. However, very little is known about this particular species or its natural products, therefore it would be interesting to characterize it further and determine how secalonic acid D fits into its chemical ecology. Interestingly, senkyunolide H was present in extracts of one isolate each of *Nigrospora sphaerica* and *Cytospora ribis*. It is most notably a plant metabolite arising from *Ligusticum chuanxiong*—an herb related to carrot that is frequently used in Traditional Chinese Medicine.[151] It is becoming increasingly common to find endophytes that produce plant secondary metabolites. This may be a result of horizontal gene

transfer between endophytes and hosts, or because endophytes and their hosts experience similar selective pressures while evolving together.[152]

One other node within the molecular network was also identified as a potentially new compound. In cluster E which contains the ellisiiamides, a node with the *m/z* 566.3914 and chemical formula of $C_{29}H_{51}N_5O_6$ (**9**) did not match any reported ellisiiamides or other cyclic pentapeptides. The likely anti-insectan activity of this new ellisiiamide should be tested in the future, along with all other reported ellisiiamides. A summary of new compounds is listed below, in Table 6.

Table 6. Unknown compounds tentatively identified from endophytic fungi as new derivatives that are targets for isolating and characterizing.

ID	Formula	*m/z*	RT (min)	Mass error (Δppm)	Structurally related to
1	$C_{17}H_{18}O_7$	335.1125 $[M+H]^+$	3.41	-0.117	Griseofulvin
2	$C_{36}H_{54}O_9$	650.4262 $[M+NH_4]^+$	3.68	-0.167	Griseofernaneoside B
3	$C_{36}H_{56}O_9$	633.3995 $[M+H]^+$	3.80	-0.363	Griseofernaneoside B
4	$C_{34}H_{52}N_4O_9$	661.3804 $[M+H]^+$	4.60	-0.507	Hirsutatin A
5	$C_{35}H_{54}N_4O_{10}$	691.3909 $[M+H]^+$	4.48	-0.044	Hirsutatin A
6	$C_{36}H_{59}N_5O_{10}$	722.4332 $[M+H]^+$	4.37	-0.414	Hirsutatin A
7	$C_{28}H_{39}NO_3$	438.3001 $[M+H]^+$	5.79	-0.298	Oxysporidinone+sambutoxin
8	$C_{29}H_{45}NO_6$	504.3321 $[M+H]^+$	5.07	0.189	Oxysporidinone+sambutoxin
9	$C_{29}H_{51}N_5O_6$	566.3914 $[M+H]^+$	4.08	0.369	Ellisiiamides

One potential shortcoming of using molecular networking as a strategy in mining datasets for new or novel compounds is that it relies on samples containing multiple compounds with similar structures to be successful. While organisms do usually produce multiple compounds with slight structural differences, this strategy is apt to overlook compounds with few or no structural "neighbours" that may be of value.

Compound screening is frequently a lengthy and difficult process which was made simpler here by the application of molecular networking. Compounds listed above in Table 6 are tentatively identified as new natural products with structural relatedness to known ones. Large-scale fermentation, extraction, isolation, and characterization by NMR are required to determine their conformations.

3.5 Conclusions and Suggestions for Future Work

The application of molecular networking to LC-MS files of Canadian fungal endophytes has identified several tentatively new compounds from fungal endophytes, including compounds related to the antifungal compound griseofulvin, two griseofernaneoside-related compounds, three hirsutatin-related compounds, a new ellisiiamide, and two new sambutoxin- and oxysporidinone-related compounds. These all have structural relatives with diverse bioactivities, which also gives the unknowns a chance of being bioactive. These compounds have the opportunity to be relevant in many areas. They may be useful in agriculture as protective agents against pests and pathogens. They may be biologically active against harmful human pathogens or may have uses as anti-cancer agents. They may be used in chemical research, perhaps as scaffolds for synthesizing new drugs. Further bioassays must be performed using isolated compounds to elicit the full scope of their uses. Molecular networking greatly simplified the process of mining a large dataset for compounds of interest, reducing the amount of time needed to pinpoint new compounds.

Future work should focus on isolating these compounds by HPLC and characterizing their structures by NMR. Furthermore, there are also still likely other interesting compounds produced by the fungal isolates included in this study. Growing these species on different types of artificial media and in different stress conditions may reveal other compound families that were not made under the growth conditions used here.

Furthermore, approximately 50 fungal isolates remain unidentified. The ITS region of some isolates was not successfully amplified by PCR, and in others, this genetic sequence was not specific enough to make an identification to the species level. Unidentified strains should be sequenced using other genetic regions, such as the translation elongation factor 1- α (TEF1α) gene to achieve better species resolution.[153] This would add to the mycological perspective, as there could be larger patterns of chemotaxonomy shared among fungal species that are not apparent with the current taxonomic identifications.

Chapter 4: Conclusions

Ilyonectria mors-panacis and *Ilyonectria robusta* are fungal pathogens that cause millions of dollars of crop loss annually in the Canadian ginseng industry.[80] Their mechanisms of pathogenicity are not wholly understood, nor are the contributions of their natural products to virulence. Because few natural products are known to be produced by these species, they were examined to determine if they produce a distinct chemical profile from other avirulent species within the same genus.

HRMS data were analyzed by PCA to ascertain a characteristic metabolomic profile of virulent species of *Ilyonectria*. PCA distinguished virulent species by their abundant production of several metabolites, including radicicol, and of some features with chemical formulas resembling radicicol. Molecular networking by GNPS determined that the formulas identified by PCA were indeed structurally related to radicicol due to their connectivity within the molecular network. These radicicol-related compounds were identified as pochonins and monocillins, which are resorcyclic acid lactones—the family to which radicicol also belongs. The antimicrobial activity of these compounds likely contributes to the survival of *Ilyonectria* in its environment, but they have not been tested for phytotoxicity. It remains to be seen if these or any other natural products made by *Ilyonectria* contribute to degradation of ginseng root, which should be examined in future studies.

The same molecular networking technique was also applied to untargeted HRMS data of 302 Canadian fungal endophytes. Fungal endophytes are the source of many natural products with use to humans, including many antimicrobial and anti-insectan compounds.[154] Extracts from fungal endophytes were subjected to molecular networking to identify new or novel compound targets for isolation and characterization. Clusters were examined that contain known bioactive

compounds, in the hopes that structurally related compounds would also possess bioactivity, making them more worthwhile targets than wholly unknown compounds. Following this process, nine new compounds were identified as targets for future studies. These include compounds related to griseofulvin, griseofernaneoside, oxysporidinone, hirsutatin A, and ellisiiamides, all of which have biological activity. Additionally, several novel sources of known compounds were identified, including potentially new sources of the bioactive compounds secalonic acid D, lovastatin, and senkyunolide H.

Across both studies, molecular networking was useful in illuminating information about fungal metabolites, whether by contextualizing the role of known compounds in the ginseng-*Ilyonectria* relationship, or by speeding up the process of identifying undiscovered compounds from Canadian fungal endophytes. The results achieved in these studies are important for characterizing the highly pressing issue of ginseng root rot, which threatens financial stability for Canadian ginseng farmers. The results from the molecular networking of fungal endophytes are important in natural product discovery and have applications in many fields. Antifungal and anti-insectan natural products can have big impacts in agriculture, protecting crops from destruction and saving money for farmers and consumers. Novel drug targets are always needed in pharmacology to address human diseases. The far-reaching potential of these natural products provides an exciting starting point for future analyses.

References

1. Demain, A. L., Importance of microbial natural products and the need to revitalize their discovery. *Journal of Industrial Microbiology & Biotechnology* **2014,** *41* (2), 185-201.
2. Blackwell, M., The Fungi: 1, 2, 3… 5.1 million species? *American Journal of Botany* **2011,** *98* (3), 426-438.
3. Kirk, P. M.; Cannon, P. F.; Minter, D. W.; Stalpers, J. A., *Dictionary of the Fungi*. 10 ed.; 2008.
4. Fleming, A., On the Antibacterial Action of Cultures of a *Penicillium*, with Special Reference to their Use in the Isolation of *B. influenzæ*. *Br J Exp Pathol* **1929,** *10* (3), 226-236.
5. Fleming, A., Penicillin. Its practical application. **1946**.
6. Rowinsky, E. K.; Donehower, R. C., Paclitaxel (Taxol). *The New England Journal of Medicine* **1995,** *332* (15), 1004-1014.
7. Cannabis extracts market data. https://www.canada.ca/en/health-canada/services/drugs-medication/cannabis/research-data/market/extracts.html (accessed June 12).
8. Turner, W. B., *Fungal Metabolites*. Academic Press Inc.: London, 1971.
9. Smedsgaard, J., Micro-scale extraction procedure for standardized screening of fungal metabolite production in cultures. *Journal of Chromatography A* **1997,** *760* (2), 264-270.
10. Croteau, R.; Kutchan, T. M.; Lewis, N. G., Natural products (secondary metabolites). In *Biochemistry and Molecular Biology of Plants*, Buchanan, B. B.; Gruissem, W.; Jones, R. L., Eds. Wiley: 2000; pp 1250-1319.
11. Degenhardt, J.; Gershenzon, J.; Baldwin, I. T.; Kessler, A., Attracting friends to feast on foes: engineering terpene emission to make crop plants more attractive to herbivore enemies. *Current Opinion in Biotechnology* **2003,** *14* (2), 169-176.
12. Madry, N.; Zocher, R.; Kleinkauf, H., Enniatin production by *Fusarium oxysporum* in chemically defined media. *European Journal of Applied Microbiology and Biotechnology* **1983,** *17* (2), 75-79.
13. Rinaldi, M. G., Use of potato flakes agar in clinical mycology. **1982,** *15* (6), 1159-1160.
14. Visagie, C. M.; Houbraken, J.; Frisvad, J. C.; Hong, S. B.; Klaassen, C. H. W.; Perrone, G.; Seifert, K. A.; Varga, J.; Yaguchi, T.; Samson, R. A., Identification and nomenclature of the genus Penicillium. *Studies in Mycology* **2014,** *78*, 343-371.
15. Smith, G., Culture media for fungi. *Journal of the Science of Food and Agriculture* **1959,** *10* (12), 674-678.
16. Scott, R. P. W., Modern liquid chromatography. *Chemical Society Reviews* **1992,** *21* (2), 137-145.
17. Schellinger, A. P.; Carr, P. W., Isocratic and gradient elution chromatography: A comparison in terms of speed, retention reproducibility and quantitation. *Journal of Chromatography A* **2006,** *1109* (2), 253-266.
18. Swartz, M. E., UPLC™: An Introduction and Review. *Journal of Liquid Chromatography & Related Technologies* **2005,** *28* (7-8), 1253-1263.
19. de Hoffman, E., Mass Spectrometry. In *Kirk-Othmer Encyclopedia of Chemical Technology*, 2005.
20. Yamashita, M.; Fenn, J. B., Electrospray ion source. Another variation on the free-jet theme. *The Journal of Physical Chemistry* **1984,** *88* (20), 4451-4459.

21. Awad, H.; Khamis, M. M.; El-Aneed, A., Mass Spectrometry, Review of the Basics: Ionization. *Applied Spectroscopy Reviews* **2015,** *50* (2), 158-175.
22. Kind, T.; Fiehn, O., Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC Bioinformatics* **2007,** *8* (1), 105.
23. Kruve, A.; Kaupmees, K., Adduct Formation in ESI/MS by Mobile Phase Additives. *Journal of The American Society for Mass Spectrometry* **2017,** *28* (5), 887-894.
24. Exactive Series Operating Manual. Scientific, T. F., Ed. Bremen, Germany, 2017; Vol. BRE0012255.
25. Kebarle, P.; Verkerk, U. H., Electrospray: From ions in solution to ions in the gas phase, what we know now. *Mass Spectrometry Reviews* **2009,** *28* (6), 898-917.
26. Wilm, M., Principles of electrospray ionization. *Molecular & Cellular Proteomics* **2011,** *10* (7), M111.009407-M111.009407.
27. Iribarne, J. V.; Thomson, B. A., On the evaporation of small ions from charged droplets. *The Journal of Chemical Physics* **1976,** *64* (6), 2287-2294.
28. de la Mora, J. F., Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism. *Analytica Chimica Acta* **2000,** *406* (1), 93-104.
29. García, M. C., The effect of the mobile phase additives on sensitivity in the analysis of peptides and proteins by high-performance liquid chromatography–electrospray mass spectrometry. *Journal of Chromatography B* **2005,** *825* (2), 111-123.
30. Chowdhury, S. K.; Katta, V.; Chait, B. T., An electrospray-ionization mass spectrometer with new features. *Rapid Communications in Mass Spectrometry* **1990,** *4* (3), 81-87.
31. Gabelica, V.; De Pauw, E., Internal energy and fragmentation of ions produced in electrospray sources. *Mass Spectrometry Reviews* **2005,** *24* (4), 566-587.
32. Michalski, A.; Damoc, E.; Hauschild, J.-P.; Lange, O.; Wieghaus, A.; Makarov, A.; Nagaraj, N.; Cox, J.; Mann, M.; Horning, S., Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Molecular Cellular Proteomics* **2011,** *10* (9).
33. Improved Sensitivity Through Enhanced Ion Transmission Using an S-Lens on the LTQ Velos Linear Ion Trap. Scientific, T. F., Ed. 2009; Vol. PSB 128, pp 1-2.
34. Olsen, J. V.; Schwartz, J. C.; Griep-Raming, J.; Nielsen, M. L.; Damoc, E.; Denisov, E.; Lange, O.; Remes, P.; Taylor, D.; Splendore, M., A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed. *Molecular and Cellular Proteomics* **2009,** *8* (12), 2759-2769.
35. Scheltema, R. A.; Hauschild, J.-P.; Lange, O.; Hornburg, D.; Denisov, E.; Damoc, E.; Kuehn, A.; Makarov, A.; Mann, M., The Q Exactive HF, a Benchtop mass spectrometer with a pre-filter, high-performance quadrupole and an ultra-high-field Orbitrap analyzer. *Mol Cell Proteomics* **2014,** *13* (12), 3698-3708.
36. Makarov, A.; Denisov, E.; Kholomeev, A.; Balschun, W.; Lange, O.; Strupat, K.; Horning, S., Performance Evaluation of a Hybrid Linear Ion Trap/Orbitrap Mass Spectrometer. *Analytical Chemistry* **2006,** *78* (7), 2113-2120.
37. Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Cooks, R. G., The Orbitrap: a new mass spectrometer. *Journal of Mass Spectrometry* **2005,** *40* (4), 430-443.
38. Benton, H. P.; Want, E. J.; Ebbels, T. M. D., Correction of mass calibration gaps in liquid chromatography–mass spectrometry metabolomics data. *Bioinformatics* **2010,** *26* (19), 2488-2489.

39. Accurate Mass. https://fiehnlab.ucdavis.edu/projects/seven-golden-rules/accurate-mass (accessed June 25).
40. Scigelova, M.; Makarov, A., Orbitrap Mass Analyzer – Overview and Applications in Proteomics. *Practical Proteomics* **2006,** *6* (S2), 16-21.
41. Makarov, A., Theory and practice of the Orbitrap mass analyzer. In *Practical Aspects of Trapped Ion Mass Spectrometry— Theory and Instrumentation*, March, R. E.; Todd, J. F., Eds. Boca Raton, FL: CRC Press: 2010; Vol. 4, pp 251-272.
42. Zubarev, R. A.; Makarov, A., Orbitrap Mass Spectrometry. *Analytical Chemistry* **2013,** *85* (11), 5288-5296.
43. Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M., Higher-energy C-trap dissociation for peptide modification analysis. *Nature Methods* **2007,** *4* (9), 709-712.
44. Mayer, P. M.; Poon, C., The mechanisms of collisional activation of ions in mass spectrometry. *Mass Spectrometry Reviews* **2009,** *28* (4), 608-639.
45. Normalized Collision Energy Technology. Scientific, T. F., Ed. Vol. PSB 104.
46. Bushee, J. L.; Argikar, U. A., An experimental approach to enhance precursor ion fragmentation for metabolite identification studies: application of dual collision cells in an orbital trap. *Rapid Communications in Mass Spectrometry* **2011,** *25* (10), 1356-1362.
47. Diedrich, J. K.; Pinto, A. F. M.; Yates, J. R., Energy Dependence of HCD on Peptide Fragmentation: Stepped Collisional Energy Finds the Sweet Spot. *Journal of the American Society for Mass Spectrometry* **2013,** *24* (11), 1690-1699.
48. Belov, M. E.; Zhang, R.; Strittmatter, E. F.; Prior, D. C.; Tang, K.; Smith, R. D., Automated Gain Control and Internal Calibration with External Ion Accumulation Capillary Liquid Chromatography-Electrospray Ionization-Fourier Transform Ion Cyclotron Resonance. *Analytical Chemistry* **2003,** *75* (16), 4195-4205.
49. Schwartz, J. C.; Zhou, X.-G.; Bier, M. E. Method and Apparatus of Increasing Dynamic Range and Sensitivity of a Mass Spectrometry. Nov. 5, 1996, 1996.
50. Kalli, A.; Smith, G. T.; Sweredoski, M. J.; Hess, S., Evaluation and optimization of mass spectrometric settings during data-dependent acquisition mode: focus on LTQ-Orbitrap mass analyzers. *J Proteome Res* **2013,** *12* (7), 3071-3086.
51. Tweeddale, H.; Notley-McRobb, L.; Ferenci, T., Effect of Slow Growth on Metabolism of *Escherichia coli*, as Revealed by Global Metabolite Pool ("Metabolome") Analysis. *Journal of Bacteriology* **1998,** *180* (19), 5109.
52. Fiehn, O., Metabolomics – the link between genotypes and phenotypes. *Plant Molecular Biology* **2002,** *48* (1), 155-171.
53. Beger, R. D., A review of applications of metabolomics in cancer. *Metabolites* **2013,** *3* (3), 552-574.
54. Brenton, A. G.; Godfrey, A. R., Accurate Mass Measurement: Terminology and Treatment of Data. *Journal of the American Society for Mass Spectrometry* **2010,** *21* (11), 1821-1835.
55. Pluskal, T.; Uehara, T.; Yanagida, M., Highly Accurate Chemical Formula Prediction Tool Utilizing High-Resolution Mass Spectra, MS/MS Fragmentation, Heuristic Rules, and Isotope Pattern Matching. *Analytical Chemistry* **2012,** *84* (10), 4396-4403.
56. Corley, D. G.; Durley, R. C., Strategies for database dereplication of natural products. *Journal of Natural Products* **1994,** *57* (11), 1484-1490.

57. Gardinassi, L. G.; Xia, J.; Safo, S. E.; Li, S., Bioinformatics Tools for the Interpretation of Metabolomics Data. *Current Pharmacology Reports* **2017,** *3* (6), 374-383.
58. Wang, Y.; Gu, M., The concept of spectral accuracy for MS. ACS Publications: 2010.
59. Pearson, K., LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **1901,** *2* (11), 559-572.
60. Lever, J.; Krzywinski, M.; Altman, N., Principal component analysis. *Nature Methods* **2017,** *14* (7), 641-642.
61. Wilcoxon, F., Individual comparisons by ranking methods. In *Breakthroughs in statistics*, Springer: 1992; pp 196-202.
62. Kruskal, W. H.; Wallis, W. A., Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association* **1952,** *47* (260), 583-621.
63. Benjamini, Y.; Hochberg, Y., Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B* **1995,** *57* (1), 289-300.
64. Hager, J. W.; Yves Le Blanc, J. C., Product ion scanning using a Q-q-Q linear ion trap (Q TRAP) mass spectrometer. *Rapid Communications in Mass Spectrometry* **2003,** *17* (10), 1056-1064.
65. Lacey, M. J.; Macdonald, C. G., Constant neutral spectrum in mass spectrometry. *Analytical Chemistry* **1979,** *51* (6), 691-695.
66. Louris, J. N.; Wright, L. G.; Cooks, R. G.; Schoen, A. E., New scan modes accessed with a hybrid mass spectrometer. *Analytical Chemistry* **1985,** *57* (14), 2918-2924.
67. Kelman, M. J.; Renaud, J. B.; Seifert, K. A.; Mack, J.; Sivagnanam, K.; Yeung, K. K. C.; Sumarah, M. W., Identification of six new *Alternaria* sulfoconjugated metabolites by high-resolution neutral loss filtering. *Rapid Communications in Mass Spectrometry* **2015,** *29* (19), 1805-1810.
68. Walsh, J. P.; Renaud, J. B.; Hoogstra, S.; McMullin, D. R.; Ibrahim, A.; Visagie, C. M.; Tanney, J. B.; Yeung, K. K. C.; Sumarah, M. W., Diagnostic fragmentation filtering for the discovery of new chaetoglobosins and cytochalasins. *Rapid Communications in Mass Spectrometry* **2019,** *33* (1), 133-139.
69. Yang, J. Y.; Sanchez, L. M.; Rath, C. M.; Liu, X.; Boudreau, P. D.; Bruns, N.; Glukhov, E.; Wodtke, A.; De Felicio, R.; Fenner, A., Molecular networking as a dereplication strategy. *Journal of Natural Products* **2013,** *76* (9), 1686-1699.
70. Wang, M.; Carver, J. J.; Phelan, V. V.; Sanchez, L. M.; Garg, N.; Peng, Y.; Nguyen, D. D.; Watrous, J.; Kapono, C. A.; Luzzatto-Knaan, T., Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nature Biotechnology* **2016,** *34* (8), 828.
71. Cooks, R. G.; Howe, I.; Williams, D. H., Structure and fragmentation mechanisms of organic ions in the mass spectrometer. *Organic Mass Spectrometry* **1969,** *2* (2), 137-156.
72. Stein, S. E.; Scott, D. R., Optimization and testing of mass spectral library search algorithms for compound identification. *Journal of the American Society for Mass Spectrometry* **1994,** *5* (9), 859-866.
73. Watrous, J.; Roach, P.; Alexandrov, T.; Heath, B. S.; Yang, J. Y.; Kersten, R. D.; van der Voort, M.; Pogliano, K.; Gross, H.; Raaijmakers, J. M., Mass spectral molecular networking of living microbial colonies. *Proceedings of the National Academy of Sciences* **2012,** *109* (26), E1743-E1752.

74. Watrous, J.; Roach, P.; Alexandrov, T.; Heath, B. S.; Yang, J. Y.; Kersten, R. D.; van der Voort, M.; Pogliano, K.; Gross, H.; Raaijmakers, J. M.; Moore, B. S.; Laskin, J.; Bandeira, N.; Dorrestein, P. C., Mass spectral molecular networking of living microbial colonies. *Proceedings of the National Academy of Sciences* **2012,** *109* (26), E1743.
75. Westerveld, S. Ginseng Production in Ontario.
76. Baeg, I.-H.; So, S.-H., The world ginseng market and the ginseng (Korea). *J Ginseng Res* **2013,** *37* (1), 1-7.
77. Baxter, M., Getting at the root of the problems facing Ontario's ginseng industry. *TVO* 2018.
78. Qi, L.-W.; Wang, C.-Z.; Yuan, C.-S., Ginsenosides from American ginseng: chemical and pharmacological diversity. *Phytochemistry* **2011,** *72* (8), 689-699.
79. Burston, C., How Canada came to dominate the global supply of ginseng. *Financial Post* 2016.
80. Seifert, K. A.; McMullen, C. R.; Yee, D.; Reeleder, R. D.; Dobinson, K. F., Molecular differentiation and detection of ginseng-adapted isolates of the root rot fungus *Cylindrocarpon destructans*. *Phytopathology* **2003,** *93* (12), 1533-1542.
81. Rahman, M.; Punja, Z. K., Factors influencing development of root rot on ginseng caused by *Cylindrocarpon destructans*. *Phytopathology* **2005,** *95* (12), 1381-1390.
82. Cabral, A.; Groenewald, J. Z.; Rego, C.; Oliveira, H.; Crous, P. W., *Cylindrocarpon* root rot: multi-gene analysis reveals novel species within the *Ilyonectria radicicola* species complex. *Mycological Progress* **2012,** *11* (3), 655-688.
83. Farh, M. E.-A.; Kim, Y.-J.; Kim, Y.-J.; Yang, D.-C., *Cylindrocarpon destructans/Ilyonectria radicicola*-species complex: Causative agent of ginseng root-rot disease and rusty symptoms. *J Ginseng Res* **2018,** *42* (1), 9-15.
84. Lu, X. H.; Zhang, X. M.; Jiao, X. L.; Hao, J. J.; Zhang, X. S.; Luo, Y.; Gao, W. W., Taxonomy of fungal complex causing red-skin root of *Panax ginseng* in China. *J Ginseng Res* **2019**.
85. Shin, J.-H.; Fu, T.; Park, K. H.; Kim, K. S., The Effect of Fungicides on Mycelial Growth and Conidial Germination of the Ginseng Root Rot Fungus, *Cylindrocarpon destructans*. *Mycobiology* **2017,** *45* (3), 220-225.
86. Westerveld, S., So you want to grow ginseng? Considerations and cautions. OMAFRA, Ed. Queen's Printer for Ontario: Ontario, 2014.
87. Reeleder, R.; Roy, R.; Capell, B., Seed and Root Rots of Ginseng (*Panax quinquefolius L*) Caused by *Cylindrocarpon destructans* and *Fusarium* spp. *J Ginseng Res* **2002,** *26* (3), 151-158.
88. Alexopoulos, C. J.; Mims, C. W., *Introductory Mycology*. 3 ed.; John Wiley & Sons: 1979.
89. Evans, G.; White, N. H., Radicicolin and radicicol, two new antibiotics produced by *Cylindrocarpon radicicola*. *Transactions of the British Mycological Society* **1966,** *49* (4), 563-567.
90. Kang, Y.; Lee, S.-H.; Lee, J., Development of a selective medium for the fungal pathogen *Cylindrocarpon destructans* using radicicol. *The Plant Pathology Journal* **2014,** *30* (4), 432.
91. Walsh, J. P.; DesRochers, N.; Renaud, J. B.; Seifert, K. A.; Yeung, K. K. C.; Sumarah, M. W., Identification of N,N′,N″-triacetylfusarinine C as a key metabolite for root rot disease virulence in American ginseng. *J Ginseng Res* **2019**.

92. Eisendle, M.; Oberegger, H.; Zadra, I.; Haas, H., The siderophore system is essential for viability of *Aspergillus nidulans*: functional analysis of two genes encoding l-ornithine N 5-monooxygenase (sidA) and a non-ribosomal peptide synthetase (sidC). *Molecular Microbiology* **2003,** *49* (2), 359-375.
93. Cossette, F.; Miller, J. D., Phytotoxic effect of deoxynivalenol and gibberella ear rot resistance of corn. *Natural Toxins* **1995,** *3* (5), 383-388.
94. Samson, R. A.; Frisvad, J. C., *Penicillium subgenus Penicillium: new taxonomic schemes and mycotoxins and other extrolites*. Centraalbureau voor Schimmelcultures: The Netherlands, 2004.
95. Dunn, W. B.; Wilson, I. D.; Nicholls, A. W.; Broadhurst, D., The importance of experimental design and QC samples in large-scale and MS-driven untargeted metabolomic studies of humans. *Bioanalysis* **2012,** *4* (18), 2249-2264.
96. DesRochers, N.; Walsh, J. P.; Renaud, J. B.; Seifert, K. A.; Yeung, K. K.-C.; Sumarah, M. W., Metabolomic Profiling of Fungal Pathogens Responsible for Root Rot in American Ginseng. *Metabolites* **2020,** *10* (1), 35.
97. Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J., A cross-platform toolkit for mass spectrometry and proteomics. *Nature Biotechnology* **2012,** *30* (10), 918.
98. Team, R. C., R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna, Austria, 2020.
99. Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G., XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Analytical Chemistry* **2006,** *78* (3), 779-787.
100. Tautenhahn, R.; Boettcher, C.; Neumann, S., Highly sensitive feature detection for high resolution LC/MS. *BMC bioinformatics* **2008,** *9* (1), 504.
101. McMillan, A.; Rulisa, S.; Sumarah, M.; Macklaim, J. M.; Renaud, J.; Bisanz, J. E.; Gloor, G. B.; Reid, G., A multi-platform metabolomics approach identifies highly specific biomarkers of bacterial diversity in the vagina of pregnant and non-pregnant women. *Scientific Reports* **2015,** *5*, 14174.
102. van den Berg, R. A.; Hoefsloot, H. C. J.; Westerhuis, J. A.; Smilde, A. K.; van der Werf, M. J., Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* **2006,** *7* (1), 142.
103. Nyamundanda, G.; Brennan, L.; Gormley, I. C., Probabilistic principal component analysis for metabolomic data. *BMC Bioinformatics* **2010,** *11* (1), 571.
104. Scrucca, L.; Fop, M.; Murphy, T. B.; Raftery, A. E., mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *The R journal* **2016,** *8* (1), 289.
105. Lê, S.; Josse, J.; Husson, F., FactoMineR: an R package for multivariate analysis. *Journal of Statistical Software* **2008,** *25* (1), 1-18.
106. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T., Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* **2003,** *13* (11), 2498-2504.
107. Evans, G.; Cartwright, J. B.; White, N. H., The production of a phytotoxin, nectrolide, by some root-surface isolates of *Cylindrocarpon radicicola*, Wr. *Plant and Soil* **1967,** *26* (2), 253-260.

108. Smith, C. A.; O'Maille, G.; Want, E. J.; Qin, C.; Trauger, S. A.; Brandon, T. R.; Custodio, D. E.; Abagyan, R.; Siuzdak, G., METLIN: a metabolite mass spectral database. *Therapeutic Drug Monitoring* **2005,** *27* (6), 747-751.
109. Hellwig, V.; Mayer-Bartschmid, A.; Müller, H.; Greif, G.; Kleymann, G.; Zitzmann, W.; Tichy, H.-V.; Stadler, M., Pochonins A− F, new antiviral and antiparasitic resorcylic acid lactones from *Pochonia chlamydosporia* var. *catenulata*. *Journal of Natural Products* **2003,** *66* (6), 829-837.
110. Aver, W. A.; Peña-Rodriguez, L., Minor metabolites of *Monocillium nordinii*. *Phytochemistry* **1987,** *26* (5), 1353-1355.
111. Shinonaga, H.; Kawamura, Y.; Ikeda, A.; Aoki, M.; Sakai, N.; Fujimoto, N.; Kawashima, A., Pochonins K–P: new radicicol analogues from *Pochonia chlamydosporia* var. *chlamydosporia* and their WNT-5A expression inhibitory activities. *Tetrahedron* **2009,** *65* (17), 3446-3453.
112. Moulin, E.; Barluenga, S.; Winssinger, N., Concise Synthesis of Pochonin A, an HSP90 Inhibitor. *Organic Letters* **2005,** *7* (25), 5637-5639.
113. Moulin, E.; Zoete, V.; Barluenga, S.; Karplus, M.; Winssinger, N., Design, Synthesis, and Biological Evaluation of HSP90 Inhibitors Based on Conformational Analysis of Radicicol and Its Analogues. *Journal of the American Chemical Society* **2005,** *127* (19), 6999-7004.
114. Whitley, D.; Goldberg, S. P.; Jordan, W. D., Heat shock proteins: a review of the molecular chaperones. *Journal of Vascular Surgery* **1999,** *29* (4), 748-751.
115. Lamoth, F.; Juvvadi, P. R.; Steinbach, W. J., Heat shock protein 90 (Hsp90): A novel antifungal target against *Aspergillus fumigatus*. *Critical Reviews in Microbiology* **2016,** *42* (2), 310-321.
116. Burgess, K. M. N.; Renaud, J. B.; McDowell, T.; Sumarah, M. W., Mechanistic insight into the biosynthesis and detoxification of fumonisin mycotoxins. *ACS Chemical Biology* **2016,** *11* (9), 2618-2625.
117. *Biological Test Method: Test for Measuring the Inhibition of Growth Using the Freshwater Macrophyte, Lemna minor*; EPS 1/RM/37; Government of Canada: Ottawa, Canada, 2007.
118. Wang, W., Literature review on duckweed toxicity testing. *Environmental Research* **1990,** *52* (1), 7-22.
119. Kusari, S.; Hertweck, C.; Spiteller, M., Chemical Ecology of Endophytic Fungi: Origins of Secondary Metabolites. *Chemistry & Biology* **2012,** *19* (7), 792-798.
120. Charbonneau, P. Debunking Endophytes.
121. Miller, J. D.; Adams, G. W. Endophyte enhanced seedlings with increased pest tolerance. 9549528, January 24, 2017, 2017.
122. Sumarah, M. W.; Miller, J. D.; Adams, G. W., Measurement of a rugulosin-producing endophyte in white spruce seedlings. *Mycologia* **2005,** *97* (4), 770-776.
123. Ibrahim, A.; Tanney, J. B.; Fei, F.; Seifert, K. A.; Cutler, G. C.; Capretta, A.; Miller, J. D.; Sumarah, M. W., Metabolomic-guided discovery of cyclic nonribosomal peptides from *Xylaria ellisii* sp. nov., a leaf and stem endophyte of *Vaccinium angustifolium*. *Scientific Reports* **2020,** *10* (1), 1-17.
124. Richardson, S. N.; Walker, A. K.; Nsiama, T. K.; McFarlane, J.; Sumarah, M. W.; Ibrahim, A.; Miller, J. D., Griseofulvin-producing *Xylaria* endophytes of *Pinus strobus*

and *Vaccinium angustifolium*: evidence for a conifer-understory species endophyte ecology. *Fungal Ecology* **2014,** *11*, 107-113.

125. Burgess, K. M. N.; Ibrahim, A.; Sørensen, D.; Sumarah, M. W., Trienylfuranol A and trienylfuranone A–B: metabolites isolated from an endophytic fungus, *Hypoxylon submoniticulosum*, in the raspberry *Rubus idaeus*. *The Journal of Antibiotics* **2017,** *70* (6), 721-725.
126. Ibrahim, A.; Sørensen, D.; Jenkins, H. A.; Ejim, L.; Capretta, A.; Sumarah, M. W., Epoxynemanione A, nemanifuranones A–F, and nemanilactones A–C, from *Nemania serpens*, an endophytic fungus isolated from Riesling grapevines. *Phytochemistry* **2017,** *140*, 16-26.
127. Pye, C. R.; Bertin, M. J.; Lokey, R. S.; Gerwick, W. H.; Linington, R. G., Retrospective analysis of natural products provides insights for future discovery trends. **2017,** *114* (22), 5601-5606.
128. Fisch, K. M., Biosynthesis of natural products by microbial iterative hybrid PKS–NRPS. *RSC Advances* **2013,** *3* (40), 18228-18247.
129. Ginn, F. M. Endophytic fungi in *Vaccinium macrocarpon* (cranberry) and *Vaccinium angustifolium* (blueberry). University of New Brunswick, 1998.
130. Schoch, C. L.; Seifert, K. A.; Huhndorf, S.; Robert, V.; Spouge, J. L.; Levesque, C. A.; Chen, W.; Consortium, F. B., Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proc Natl Acad Sci U S A* **2012,** *109* (16), 6241-6246.
131. Roullier, C.; Bertrand, S.; Blanchet, E.; Peigné, M.; Robiou du Pont, T.; Guitton, Y.; Pouchus, Y. F.; Grovel, O., Time dependency of chemodiversity and biosynthetic pathways: an LC-MS metabolomic study of marine-sourced *Penicillium*. *Marine Drugs* **2016,** *14* (5), 103.
132. Blank, H.; Roth, F. J., Jr.; Bruce, W. W.; Engel, M. F.; Smith, J. G., Jr.; Zaias, N., The Treatment of Dermatomycoses with Orally Administered Griseofulvin. *Archives of Dermatology* **1982,** *118* (10), 827-834.
133. Finkelstein, E.; Amichai, B.; Grunwald, M. H., Griseofulvin and its uses. *International Journal of Antimicrobial Agents* **1996,** *6* (4), 189-194.
134. Ibrahim, A. Microbial Secondary Metabolomics for Natural Product Discovery. McMaster University, Hamilton, Ontario, Canada, 2017.
135. Park, J.-H.; Choi, G. J.; Lee, H. B.; Kim, K. M.; Jung, H. S.; Lee, S. W.; Jang, K. S.; Cho, K. Y.; Kim, J. C., Griseofulvin from *Xylaria* sp. strain F0010, an endophytic fungus of *Abies holophylla* and its antifungal activity against plant pathogenic fungi. *Journal of Microbiology and Biotechnology* **2005,** *15* (1), 112-117.
136. Brian, P. W.; Curtis, P. J.; Hemming, H. G., A substance causing abnormal development of fungal hyphae produced by *Penicillium janczewskii* zal.: I. Biological assay, production and isolation of 'curling factor'. *Transactions of the British Mycological Society* **1946,** *29* (3), 173-187.
137. Shang, Z.; Li, X. M.; Li, C. S.; Wang, B. G., Diverse Secondary Metabolites Produced by Marine-Derived Fungus *Nigrospora* sp. MA75 on Various Culture Media. *Chemistry & Biodiversity* **2012,** *9* (7), 1338-1348.
138. Hamed, A. I.; Masullo, M.; Pecio, L.; Gallotta, D.; Mahalel, U. A.; Pawelec, S.; Stochmal, A.; Piacente, S., Unusual fernane and gammacerane glycosides from the aerial parts of *Spergula fallax*. *Journal of Natural Products* **2014,** *77* (3), 657-662.

139. Isaka, M.; Palasarn, S.; Sriklung, K.; Kocharin, K., Cyclohexadepsipeptides from the Insect Pathogenic Fungus *Hirsutella nivea* BCC 2594. *Journal of Natural Products* **2005,** *68* (11), 1680-1682.
140. Breinholt, J.; Ludvigsen, S.; Rassing, B. R.; Rosendahl, C. N.; Nielsen, S. E.; Olsen, C. E., Oxysporidinone: A Novel, Antifungal N-Methyl-4-hydroxy-2-pyridone from *Fusarium oxysporum*. *Journal of Natural Products* **1997,** *60* (1), 33-35.
141. Kim, J.-C.; Lee, Y.-W.; Yu, S.-H., Sambutoxin-producing isolates of *Fusarium* species and occurrence of sambutoxin in rotten potato tubers. *Applied and Environmental Microbiology* **1995,** *61* (10), 3750-3751.
142. Kim, J.-C.; Lee, Y.-W., Sambutoxin, a new mycotoxin produced by toxic *Fusarium* isolates obtained from rotted potato tubers. *Applied and Environmental Microbiology* **1994,** *60* (12), 4380-4386.
143. Jessen, H. J.; Gademann, K., 4-Hydroxy-2-pyridone alkaloids: Structures and synthetic approaches. *Natural Product Reports* **2010,** *27* (8), 1168-1185.
144. Tsuchinari, M.; Shimanuki, K.; Hiramatsu, F.; Murayama, T.; Koseki, T.; Shiono, Y., Fusapyridons A and B, novel pyridone alkaloids from an endophytic fungus, *Fusarium* sp. YG-45. *Zeitschrift für Naturforschung B* **2007,** *62* (9), 1203-1207.
145. Cardani, C.; Casnati, G.; Piozzi, F.; Quilico, A., The Constitution of Echinulin. *Tetrahedron Letters* **1959,** *1* (16), 1-8.
146. Endo, A.; Kuroda, M.; Tsujita, Y., ML-236A, ML-236B, and ML-236C, new inhibitors of cholesterogenesis produced by *Penicillium citrinum*. *The Journal of antibiotics* **1976,** *29* (12), 1346-1348.
147. Arsenijevic, M.; Boric, B.; Draganic, M.; Spica, G.; Aleksic, G., Cultural characteristics and pathogenicity of *Seimatosporium lichenicola* (Corda) Shoemaker et Müller isolated from blackberry (*Rubus fruticosus* L.) plants in Yugoslavia. *Journal of Plant Diseases and Protection* **1999**, 353-362.
148. Steyn, P. S., The isolation, structure and absolute configuration of secalonic acid D, the toxic metabolite of *Penicillium oxalicum*. *Tetrahedron* **1970,** *26* (1), 51-57.
149. Andersen, R.; Buechi, G.; Kobbe, B.; Demain, A. L., Secalonic acids D and F are toxic metabolites of *Aspergillus aculeatus*. *The Journal of Organic Chemistry* **1977,** *42* (2), 352-353.
150. De Vrije, T.; Antoine, N.; Buitelaar, R.; Bruckner, S.; Dissevelt, M.; Durand, A.; Gerlagh, M.; Jones, E.; Lüth, P.; Oostra, J., The fungal biocontrol agent *Coniothyrium minitans:* production by solid-state fermentation, application and marketing. *Applied Microbiology and Biotechnology* **2001,** *56* (1-2), 58-68.
151. Li, S.-L.; Yan, R.; Tam, Y.-K.; Lin, G., Post-Harvest Alteration of the Main Chemical Ingredients in Ligusticum chuanxiong Hort. (*Rhizoma Chuanxiong*). *Chemical Pharmaceutical Bulletin* **2007,** *55* (1), 140-144.
152. Aly, A. H.; Debbab, A.; Proksch, P., Fungal endophytes–secret producers of bioactive plant metabolites. *Die Pharmazie-An International Journal of Pharmaceutical Sciences* **2013,** *68* (7), 499-505.
153. Stielow, J. B.; Lévesque, C. A.; Seifert, K. A.; Meyer, W.; Iriny, L.; Smits, D.; Renfurm, R.; Verkley, G. J. M.; Groenewald, M.; Chaduli, D.; Lomascolo, A.; Welti, S.; Lesage-Meessen, L.; Favel, A.; Al-Hatmi, A. M. S.; Damm, U.; Yilmaz, N.; Houbraken, J.; Lombard, L.; Quaedvlieg, W.; Binder, M.; Vaas, L. A. I.; Vu, D.; Yurkov, A.; Begerow, D.; Roehl, O.; Guerreiro, M.; Fonseca, A.; Samerpitak, K.; van

Diepeningen, A. D.; Dolatabadi, S.; Moreno, L. F.; Casaregola, S.; Mallet, S.; Jacques, N.; Roscini, L.; Egidi, E.; Bizet, C.; Garcia-Hermoso, D.; Martín, M. P.; Deng, S.; Groenewald, J. Z.; Boekhout, T.; de Beer, Z. W.; Barnes, I.; Duong, T. A.; Wingfield, M. J.; de Hoog, G. S.; Crous, P. W.; Lewis, C. T.; Hambleton, S.; Moussa, T. A. A.; Al-Zahrani, H. S.; Almaghrabi, O. A.; Louis-Seize, G.; Assabgui, R.; McCormick, W.; Omer, G.; Dukik, K.; Cardinali, G.; Eberhardt, U.; de Vries, M.; Robert, V., One fungus, which genes? Development and assessment of universal primers for potential secondary fungal DNA barcodes. *Persoonia* **2015,** *35*, 242-263.

154. Schulz, B.; Boyle, C.; Draeger, S.; Römmert, A.-K.; Krohn, K., Endophytic fungi: a source of novel biologically active secondary metabolites. *Mycological Research* **2002,** *106* (9), 996-1004.
155. Smith, C. A.; Tautenhahn, R.; Neumann, S.; Benton, H. P.; Conley, C.; Rainer, J., Package "xcms". 3.4.4 ed.; 2019.
156. McMillan, A. xcms Workflow. https://github.com/amcmil/Metabolomics-LC-MS/blob/master/xcms_workflow.R.
157. Chong, J.; Yamamoto, M.; Xia, J., MetaboAnalystR 2.0: From Raw Spectra to Biological Insights. *Metabolites* **2019,** *9* (3), 57.
158. Dunn, W. B.; Broadhurst, D.; Brown, M.; Baker, P. N.; Redman, C. W. G.; Kenny, L. C.; Kell, D. B., Metabolic profiling of serum using Ultra Performance Liquid Chromatography and the LTQ-Orbitrap mass spectrometry system. *Journal of Chromatography B* **2008,** *871* (2), 288-298.

Appendix A: Chapter 1 Supplementary

Table A1. Select *xcms* package functions and parameters. Adapted from the *xcms* users' manual.[155, 156]

***xcms* Function**	**Function arguments and explanations**	
xcmsSet	method	Specifies the method of detecting LC-MS features. "centWave" method is used in these analyses[100]
	prefilter (k, I)	Specifies numeric values for prefiltering data while determining regions of interest. Mass traces are only kept if they have at least k peaks with I intensity
	ppm	Defines the maximum *m/z* tolerance between peaks that may be considered the same. A value of 5 is recommended for Orbitrap Q-Exactive instruments[157]
	snthresh	Sets the signal to noise ratio cut-off. Although there is no S/N ratio in Orbitrap instruments, this parameter significantly affects the number of detected features. Should be adjusted according to individual datasets[158]
	peakwidth	A time range given as (min, max) in seconds, that states the expected peak width of chromatographic peaks[157]
	noise	Sets the minimum intensity a centroid peak must have to be included in the final peaklist
group	bw	Standard deviation or half width at half maximum of smoothing kernel. Strongly influences the number of detected features
	minfrac	Sets the minimum number of samples that must contain a feature for it to be considered real
	mzwid	The width of overlapping *m/z* slices used to make peak density chromatograms and group peaks across samples
fillPeaks	xcmsSet object	Fills the peaklist with peaks from raw files based on features that were compiled in previous steps

Appendix B: Chapter 2 Supplementary

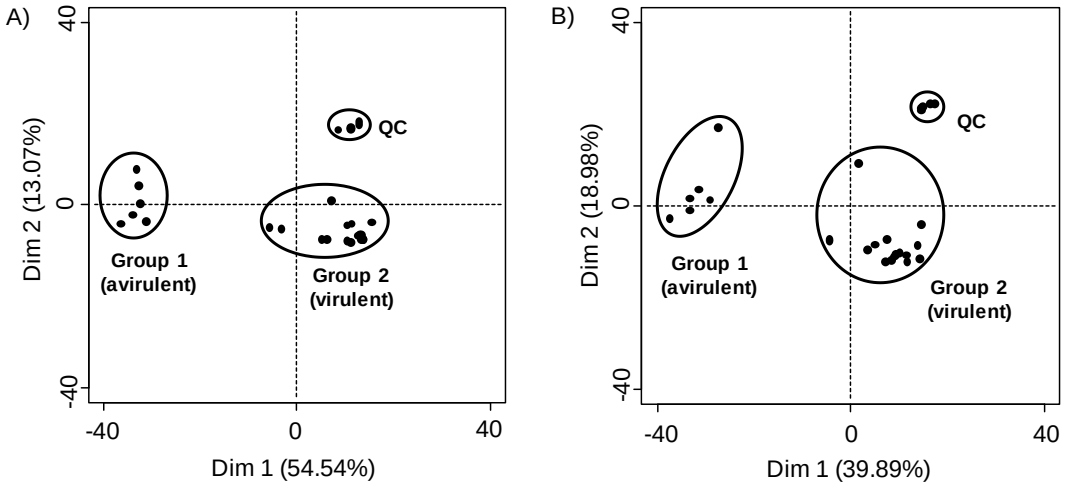


Figure B1. Principal Component Analysis plots for metabolites produced by *Ilyonectria* spp. in A) positive ionization mode, and B) negative ionization mode. Both plots were produced with LC-HRMS data processed with *xcms*. Quality control samples were included to assess instrument drift over time. (Reproduced from *Metabolites*, open access journal).[96]

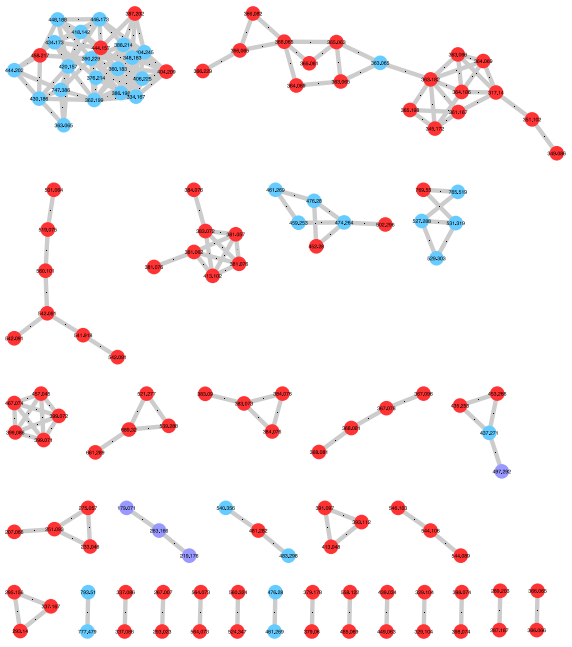


Figure B2. Molecular network of negative mode tandem MS data from strains of *Ilyonectria* and *Neonectria*. Red: produced by virulent strains; violet: produced by avirulent strains; blue: produced by both virulent and avirulent strains.

Table B1. Twenty most abundant metabolites by peak area in positive ionization mode, produced by virulent strains of *Ilyonectria*. (Significantly different from avirulent strains by a Kruskal-Wallis test with Benjamini-Hochberg *p*-value correction, $p < 0.05$). (Reproduced from *Metabolites*, open access journal).[96]

m/z $[M+H]^+$	Rt (min)	Formula	*p*-value
205.1951	4.43	$C_{15}H_{24}$	0.002
205.1951	3.55	$C_{15}H_{24}$	0.011
221.1899	4.47	$C_{15}H_{24}O$	0.003
233.0807	2.77	$C_{13}H_{12}O_4$	0.002
235.0964	2.68	$C_{13}H_{14}O_4$	0.005
253.1796	4.38	$C_{15}H_{24}O_3$	0.002
333.133	3.17	$C_{18}H_{20}O_6$	0.002
349.0834	2.99	$C_{18}H_{17}ClO_5$	0.002
351.0992	3.67	$C_{18}H_{19}ClO_5$	0.002
365.0785	3.19	$C_{18}H_{17}ClO_6$	0.002
365.0786	2.65	$C_{18}H_{17}ClO_6$	0.002
367.0937	3.32	$C_{18}H_{19}ClO_6$	0.002
383.089	2.92	$C_{18}H_{19}ClO_7$	0.002
383.0891	2.57	$C_{18}H_{19}ClO_7$	0.002
388.2115	3.02	$C_{22}H_{29}NO_5$	0.011
406.105	3.19	$C_{20}H_{21}NClO_6$	0.002
496.3631	3.74	$C_{24}H_{45}N_7O_4$	0.031
498.3787	3.93	$C_{24}H_{47}N_7O_4$	0.049
500.3943	4.17	$C_{24}H_{49}N_7O_4$	0.031
518.3241	3.62	$C_{29}H_{39}N_7O_2$	0.002

Table B2. Twenty most abundant metabolites by peak area in negative ionization mode, produced by virulent strains of *Ilyonectria*. (Significantly different from avirulent strains by a Kruskal-Wallis test with Benjamini-Hochberg *p*-value correction, $p < 0.05$). (Reproduced from *metabolites*, open access journal).[96]

m/z [M-H]-	Rt (min)	Formula	*p*-value
201.1124	2.45	$C_{10}H_{18}O_4$	0.001
259.0610	3.07	$C_{14}H_{12}O_5$	0.002
275.0561	2.94	$C_{29}H_{51}NO_8$	0.002
275.1864	2.72	$C_{14}H_{28}O_5$	0.024
349.0847	3.67	$C_{18}H_{19}ClO_5$	0.001
363.0641	3.19	$C_{18}H_{17}ClO_6$	0.001
363.0641	3.65	$C_{18}H_{17}ClO_6$	0.001
365.0795	2.99	$C_{18}H_{19}ClO_6$	0.001
367.0951	3.34	$C_{18}H_{21}ClO_6$	0.001
379.0588	3.18	$C_{18}H_{17}ClO_7$	0.001
379.0590	2.95	$C_{18}H_{17}ClO_7$	0.001
381.0746	2.74	$C_{18}H_{19}ClO_7$	0.001
383.0899	2.87	$C_{18}H_{21}ClO_7$	0.001
383.0903	3.21	$C_{18}H_{21}ClO_7$	0.001
393.1111	4.08	$C_{20}H_{23}ClO_6$	0.001
397.0694	2.74	$C_{18}H_{19}ClO_8$	0.001
399.0851	2.65	$C_{18}H_{21}ClO_8$	0.001
526.0942	2.88	$C_{23}H_{24}NSClO_9$	0.001
540.3541	3.74	$C_{29}H_{51}NO_8$	0.024
542.3698	3.93	$C_{29}H_{53}NO_8$	0.030

Appendix C: Chapter 3 Supplementary

Table C1. List of 302 endophyte isolates used for GNPS analysis, with tentative species identification based on ITS sequencing, geographic origin, and host organism listed where available.

Endophyte ID	Tentative species ID	Origin	Host Organism
E-002	*Epicoccum* sp.	Bala, ON	Cranberry
E-003	Unknown fungal endophyte	Bala, ON	Cranberry
E-004	*Botrytis* sp.	Bala, ON	Cranberry
E-005	*Epicoccum* sp.	Bala, ON	Cranberry
E-006	*Nigrospora* cf. *sphaerica*	Bala, ON	Cranberry
E-007	*Varicosporium delicatum*	Bala, ON	Cranberry
E-008	*Nemania serpens*	Bala, ON	Cranberry
E-009	*Venturia* sp.	Bala, ON	Cranberry
E-010	cf. *Protoventuria barriae*	Bala, ON	Cranberry
E-011	*Seimatosporium lichenicola*	Debert, NS	Lowbush blueberry
E-012	Unknown fungal endophyte	Debert, NS	Lowbush blueberry
E-014	*Godronia cassandrae*	Portapique, NS	Lowbush blueberry
E-015	*Paramycosphaerella watsoniae*	Portapique, NS	Lowbush blueberry
E-016	*Didymella* sp.	Portapique, NS	Lowbush blueberry
E-017	*Wettsteinina* cf. *macrotheca*	Portapique, NS	Lowbush blueberry
E-018	cf. *Spondycladiopsis cupulicola*	Portapique, NS	Lowbush blueberry
E-019	*Herpotrichia* cf. *juniperi*	Portapique, NS	Lowbush blueberry
E-020	*Guignardia vaccinii*	Portapique, NS	Lowbush blueberry
E-021	cf. *Rhizosphaera* sp.	Portapique, NS	Lowbush blueberry
E-022	cf. *Rhizosphaera* sp.	Mt. Thom, NS	Lowbush blueberry
E-023	cf. *Rhizosphaera* sp.	Mt. Thom, NS	Lowbush blueberry
E-024	cf. *Leptodontidium* sp.	Mt. Thom, NS	Lowbush blueberry
E-025	cf. *Rhizosphaera* sp.	Mt. Thom, NS	Lowbush blueberry
E-026	*Coniochaeta tetraspora*	Mt. Thom, NS	Lowbush blueberry
E-027	Unknown fungal endophyte	Mt. Thom, NS	Lowbush blueberry
E-028	*Xylaria* sp.	Mt. Thom, NS	Lowbush blueberry
E-029	Unknown fungal endophyte	Mt. Thom, NS	Lowbush blueberry

E-030	*Phialocephala* sp.	Mt. Thom, NS	Lowbush blueberry
E-031	*Phialocephala* sp.	Mt. Thom, NS	Lowbush blueberry
E-032	cf. *Godronia cassandrae*	Mt. Thom, NS	Lowbush blueberry
E-033	cf. *Rhizosphaera* sp.	Bala, ON	Cranberry
E-034	*Coprinellus micaceus*	Mt. Thom, NS	Lowbush blueberry
E-035	cf. *Leptodontidium* sp.	Mt. Thom, NS	Lowbush blueberry
E-036	*Godronia cassandrae*	Mt. Thom, NS	Lowbush blueberry
E-038	cf. *Rhizosphaera* sp.	Bala, ON	Cranberry
E-039	*Mollisia* sp.	Mt. Thom, NS	Lowbush blueberry
E-040	cf. *Mollisia nigrescens*	Simcoe, ON	Highbush blueberry
E-041	Unknown fungal endophyte	Mt. Thom, NS	Lowbush blueberry
E-042	*Pyrenopeziza* sp.	Portapique, NS	Lowbush blueberry
E-043	*Lachnum* sp.	Simcoe, ON	Highbush blueberry
E-044	*Phialocephala* sp.	Mt. Thom, NS	Lowbush blueberry
E-045	Unknown fungal endophyte	Mt. Thom, NS	Lowbush blueberry
E-046	*Godronia* cf. *cassandrae*	Portapique, NS	Lowbush blueberry
E-047	*Paramycosphaerella watsoniae*	Portapique, ON	Lowbush blueberry
E-048	Unknown fungal endophyte	Parks Blueberries, ON	Highbush blueberry
E-049	*Phyllosticta* sp.	Mt. Thom, NS	Lowbush blueberry
E-050	*Nemania serpens*	Simcoe, ON	Highbush blueberry
E-051	*Coniochaeta tetraspora*	Portapique, ON	Lowbush blueberry
E-052	cf. *Mycosphaerella* sp.	Portapique, ON	Lowbush blueberry
E-053	cf. *Leptodontidium* sp.	Mt. Thom, NS	Lowbush blueberry
E-054	*Phyllosticta* sp.	P.E.I.	Lowbush blueberry
E-055	Unknown fungal endophyte	Bala, ON	Cranberry
E-056	Unknown fungal endophyte	Bala, ON	Cranberry
E-058	Unknown fungal endophyte	Jordan Station, ON	Grape
E-059	*Setomelanomma* sp.	Bala, ON	Cranberry
E-060	Unknown fungal endophyte	Mt. Thom, NS	Lowbush blueberry
E-061	Unknown fungal endophyte	Portapique, NS	Lowbush blueberry
E-062	*Diaporthe* cf. *caulivora*	Simcoe, ON	Highbush blueberry
E-063	*Mollisia* sp.	Portapique, NS	Lowbush blueberry
E-064	Unknown fungal endophyte	Jordan Station, ON	Highbush blueberry
E-065	cf. *Seimatosporium lichenicola*	Jordan Station, ON	Highbush blueberry
E-066	*Didymella* sp.	Parks Blueberries, ON	Highbush blueberry

E-067	*Geastrumia polystigmatis*	Mt. Thom, NS	Lowbush blueberry
E-068	Unknown fungal endophyte	Bala, ON	Cranberry
E-069	Unknown fungal endophyte	Debert, NS	Lowbush blueberry
E-070	Unknown fungal endophyte	Debert, NS	Lowbush blueberry
E-071	*Godronia* cf. *cassandrae*	Portapique, NS	Lowbush blueberry
E-072	*Mollisia* sp.	Bala, ON	Cranberry
E-073	*Sphaerulina* cf. *rhabdoclinis*	Debert, NS	Lowbush blueberry
E-075	*Xylaria* cf. *castorea*	Jordan Station, ON	Grape
E-076	*Lachnum* sp.	Bala, ON	Cranberry
E-077	*Lophiostoma* sp.	Debert, NS	Lowbush blueberry
E-078	Unknown fungal endophyte	Jordan Station, ON	Highbush blueberry
E-079	*Lophiostoma* sp.	Mt. Thom, NS	Lowbush blueberry
E-081	cf. *Rhizosphaera* sp.	Debert, NS	Lowbush blueberry
E-082	*Hypoxylon submonticulosum*	Jordan Station, ON	Raspberry
E-083	From Rhytismataceae family	Portapique, NS	Lowbush blueberry
E-084	*Hypoxylon submonticulosum*	Jordan Station, ON	Pear
E-086	Unknown fungal endophyte	Portapique, NS	Lowbush blueberry
E-087	*Coniochaeta tetraspora*	Portapique, NS	Lowbush blueberry
E-088	*Phyllosticta* cf. *rubra*	Mt. Thom, NS	Lowbush blueberry
E-089	*Godronia* cf. *cassandrae*	Portapique, NS	Lowbush blueberry
E-090	Unknown fungal endophyte	Portapique, NS	Lowbush blueberry
E-091	*Sphaerulina* cf. *rhabdoclinis*	Portapique, NS	Lowbush blueberry
E-092	*Trichaptum abietinum*	Debert, NS	Lowbush blueberry
E-093	Unknown fungal endophyte	Bala, ON	Cranberry
E-094	Unknown fungal endophyte	Debert, NS	Lowbush blueberry
E-095	Unknown fungal endophyte	Mt. Thom, NS	Lowbush blueberry
E-096	Unknown fungal endophyte	Portapique, NS	Lowbush blueberry
E-097	*Microdochium neoqueenslandicum*	Rawdon, NS	Lowbush blueberry
E-099	Unknown fungal endophyte	Jordan Station, ON	Grape
E-100	Unknown fungal endophyte	Mt, Thom, NS	Lowbush blueberry
E-101	*Creosphaeria sassafras*	Jordan Station, ON	Grape
E-102	cf. *Leptodontidium* sp.	Portapique, NS	Lowbush blueberry
E-103	*Nemania serpens*	Rawdon, NS	Highbush blueberry
E-104	*Xylaria ellisii*	Rawdon, NS	Highbush blueberry

E-105	Unknown fungal endophyte	Mt. Thom, NS	Lowbush blueberry
E-106	Unknown fungal endophyte	Mt. Thom, NS	Lowbush blueberry
E-107	*Xylaria ellisii*	Mt. Thom, NS	Lowbush blueberry
E-108	Unknown fungal endophyte	Mt. Thom, NS	Lowbush blueberry
E-109	Unknown fungal endophyte	Mt. Thom, NS	Lowbush blueberry
E-110	Unknown fungal endophyte	Mt. Thom, NS	Lowbush blueberry
E-111	From Rhytismataceae family	Debert, NS	Lowbush blueberry
E-112	*Xylaria ellisii*	Mt. Thom, NS	Lowbush blueberry
E-113	cf. *Leptodontidium* sp.	Mt. Thom, NS	Lowbush blueberry
E-114	*Xylaria ellisii*	Debert, NS	Lowbush blueberry
E-115	*Xylaria ellisii*	Rawdon, NS	Highbush blueberry
E-116	*Xylaria ellisii*	Portapique, NS	Lowbush blueberry
E-117	*Pilidium* sp.	Portapique, NS	Lowbush blueberry
E-118	*Pilidium* sp.	Portapique, NS	Lowbush blueberry
E-119	*Mollisia* cf. *melaleuca*	Debert, NS	Lowbush blueberry
E-120	Unknown fungal endophyte	Portapique, NS	Lowbush blueberry
E-121	PCR unsuccessful	Portapique, NS	Lowbush blueberry
E-122	*Alternaria* sp.	Mt. Thom, NS	Lowbush blueberry
E-123	*Xylaria* sp.	Rawdon, NS	Highbush blueberry
E-124	cf. *Mollisia nigrescens*	Portapique, NS	Lowbush blueberry
E-125	*Alternaria* sp.	Jordan Station, ON	Raspberry
E-129	*Godronia* cf. *cassandrae*	Portapique, NS	Lowbush blueberry
E-130	From Rhytismataceae family	Debert, NS	Lowbush blueberry
E-131	PCR unsuccessful	Jordan Station, ON	Highbush blueberry
E-133	cf. *Devriesia* sp.	Mt. Thom, NS	Lowbush blueberry
E-135	*Sphaerulina* sp.	Mt. Thom, NS	Lowbush blueberry
E-138	*Xylaria ellisii*	Rawdon, NS	Highbush blueberry
E-140	*Nemania serpens*	Jordan Station, ON	Grape
E-141	Unknown fungal endophyte	Mt. Thom, NS	Lowbush blueberry
E-142	*Xylaria ellisii*	Portapique, NS	Lowbush blueberry
E-143	*Xylaria ellisii*	Portapique, NS	Lowbush blueberry
E-144	*Wettsteinina* cf. *mirabilis*	Rawdon, NS	Highbush blueberry
E-145	*Alternaria* sp.	Jordan Station, ON	Raspberry
E-148	cf. *Nemania* sp.	Jordan Station, ON	Grape
E-150	*Xylaria ellisii*	Mt. Thom, NS	Lowbush blueberry

E-152	cf. *Rhizosphaera* sp.	Portapique, NS	Lowbush blueberry
E-153	*Xylaria ellisii*	Portapique, NS	Lowbush blueberry
E-154	cf. *Mollisia nigrescens*	Portapique, NS	Lowbush blueberry
E-155	cf. *Leptodontidium*	Portapique, NS	Lowbush blueberry
E-156	*Nemania diffusa*	Jordan Station, ON	Highbush blueberry
E-157	*Xylaria ellisii*	Portapique, NS	Lowbush blueberry
E-158	Unknown fungal endophyte	Portapique, NS	Lowbush blueberry
E-159	cf. *Nemania* sp.	Portapique, NS	Lowbush blueberry
E-160	cf. *Seimatosporium lichenicola*	Portapique, NS	Lowbush blueberry
E-161	Unknown fungal endophyte	Portapique, NS	Lowbush blueberry
E-162	*Xylaria ellisii*	Rawdon, NS	Highbush blueberry
E-163	Unknown fungal endophyte	Rawdon, NS	Highbush blueberry
E-164	*Xylaria ellisii*	Debert, NS	Lowbush blueberry
E-165	*Phaeoacremonium* sp.	Rawdon, NS	Highbush blueberry
E-167	*Xylaria* sp.	Debert, NS	Lowbush blueberry
E-168	Unknown fungal endophyte	Debert, NS	Lowbush blueberry
E-169	*Alternaria* sp.	Jordan Station, ON	Raspberry
E-170	*Xylaria ellisii*	Mt. Thom, NS	Lowbush blueberry
E-171	*Alternaria infectoria*	Jordan Station, ON	Highbush blueberry
E-172	cf. *Anthostomella* sp.	Jordan Station, ON	Grape
E-173	*Nemania serpens*	Jordan Station, ON	Raspberry
E-174	*Phomopsis* cf. *vaccinii*	Mt. Thom, NS	Blueberry
E-175	*Phyllosticta* cf. *pyrolae*	Debert, NS	Blueberry
E-176	*Phyllosticta* cf. *rubra*	Portapique, NS	Blueberry
E-177	*Paraphaeosphaeria neglecta*	Mt. Thom, NS	Blueberry
E-178	*Fusarium tricinctum**	Mt. Thom, NS	Blueberry
E-179	*Sphaerulina* cf. *rhabdoclinis*	Debert, NS	Blueberry
E-180	*Phomopsis* cf. *vaccinii*	Debert, NS	Blueberry
E-181	*Epicoccum nigrum*	Debert, NS	Blueberry
E-182	cf. *Paraphaeosphaeria neglecta*	Mt. Thom, NS	Blueberry
E-183	*Godronia* cf. *cassandrae*	Portapique, NS	Blueberry
E-184	*Xylaria cubensis*	Simcoe, ON	Grape
E-188	cf. *Rhizosphaera* sp.	Mt. Thom, NS	Blueberry
E-189	*Paraphaeosphaeria neglecta*	Mt. Thom, NS	Blueberry
E-190	*Sphaerulina* cf. *rhabdoclinis*	Mt. Thom, NS	Blueberry

E-191	cf. *Rhizosphaera* sp.	Mt. Thom, NS	Blueberry
E-192	*Sphaerulina* cf. *rhabdoclinis*	Mt. Thom, NS	Blueberry
E-193	cf. *Rhizosphaera* sp.	Mt. Thom, NS	Blueberry
E-194	*Nemania* cf. *serpens*	Jordan Station, ON	Grape
E-196	cf. *Mollisia nigrescens*	Portapique, NS	Blueberry
E-197	*Cryptosporella femoralis*	Portapique, NS	Blueberry
E-198	*Phomopsis* cf. *vaccinii*	Mt. Thom, NS	Blueberry
E-199	*Kretzschmaria* cf. *deusta*	Jordan Station, ON	Blueberry
E-200	*Fusarium tricinctum*	Mt. Thom, NS	Blueberry
E-201	cf. *Claussenomyces* sp.	Debert, NS	Blueberry
E-202	*Nemania serpens*	Jordan Station, ON	Grape
E-203	*Coniothyrium ferrarisianum*	Jordan Station, ON	Grape
E-204	*Phyllosticta* cf. *pyrolae*	Mt. Thom, NS	Blueberry
E-205	*Godronia* cf. *cassandrae*	Mt. Thom, NS	Blueberry
E-206	*Xylaria ellisii*	Portapique, NS	Blueberry
E-207	cf. *Seimatosporium lichenicola*	New Brunswick, CA	Blueberry
E-208	*Xylaria ellisii*	New Brunswick, CA	Blueberry
E-209	*Godronia* cf. *cassandrae*	New Brunswick, CA	Blueberry
E-210	*Godronia* cf. *cassandrae*	New Brunswick, CA	Blueberry
E-211	*Godronia* cf. *cassandrae*	New Brunswick, CA	Blueberry
E-212	From Rhytismataceae family	New Brunswick, CA	Blueberry
E-213	*Godronia* cf. *cassandrae*	New Brunswick, CA	Blueberry
E-214	From Rhytismataceae family	New Brunswick, CA	Blueberry
E-215	*Ramularia* cf. *nyssicola*	New Brunswick, CA	Blueberry
E-216	*Xylaria ellisii*	New Brunswick, CA	Blueberry
E-217	cf. *Mollisia nigrescens*	Portapique, NS	Blueberry
E-218	*Mollisia* sp.	New Brunswick, CA	Blueberry
E-219	*Stemphylium globuliferum*	Debert, NS	Blueberry
E-220	From Sordariomycetes class	Jordan Station, ON	Grape
E-221	cf. *Leptodontidium* sp.	Portapique, NS	Blueberry
E-222	Unknown fungal endophyte	Mt. Thom, NS	Blueberry
E-223	*Nemania serpens*	Debert, NS	Blueberry
E-224	*Paraphaeosphaeria neglecta*	Mt. Thom, NS	Blueberry
E-225	*Phyllosticta* cf. *rhabdoclinis*	New Brunswick, CA	Blueberry
E-226	*Xylaria ellisii*	New Brunswick, CA	Blueberry

E-227	From Rhytismataceae family	New Brunswick, CA	Blueberry
E-228	*Paraphaeosphaeria neglecta*	Debert, NS	Blueberry
E-229	*Phomopsis* cf. *vaccinii*	Portapique, NS	Blueberry
E-230	*Sphaerulina* cf. *rhabdoclinis*	Mt. Thom, NS	Blueberry
E-231	Unknown fungal endophyte	Mt. Thom, NS	Blueberry
E-232	cf. *Mycosphaerella* sp.	Mt. Thom, NS	Blueberry
E-233	cf. *Seimatosporium lichenicola*	New Brunswick, CA	Blueberry
E-234	PCR unsuccessful	New Brunswick, CA	Blueberry
E-235	cf. *Seimatosporium lichenicola*	Portapique, NS	Blueberry
E-236	Unknown fungal endophyte	Jordan Station, ON	Grape
E-237	*Colletotrichum* cf. *truncatum*	Mt. Thom, NS	Blueberry
E-238	Unknown fungal endophyte	New Brunswick, CA	Blueberry
E-239	From Rhytismataceae family	New Brunswick, CA	Blueberry
E-240	cf. *Claussenomyces* sp.	New Brunswick, CA	Blueberry
E-241	*Xylaria ellisii*	New Brunswick, CA	Blueberry
E-242	PCR unsuccessful	New Brunswick, CA	Blueberry
E-243	*Xylaria ellisii*	Mt. Thom, NS	Blueberry
E-244	*Xylaria ellisii*	Mt. Thom, NS	Blueberry
E-245	*Godronia* cf. *cassandrae*	Mt. Thom, NS	Blueberry
E-246	*Godronia* cf. *cassandrae*	Mt. Thom, NS	Blueberry
E-247	cf. *Claussenomyces* sp.	Mt. Thom, NS	Blueberry
E-248	*Ramularia* cf. *nyssicola*	New Brunswick, CA	Blueberry
E-249	*Xylaria* cf. *castorea*	Jordan Station, ON	Blueberry
E-250	*Hypoxylon submonticulosum*	Jordan Station, ON	Grape
E-251	*Hypoxylon submonticulosum*	Jordan Station, ON	Raspberry
E-252	*Ramularia* cf. *bellunensis*	New Brunswick, CA	Blueberry
E-253	*Ramularia* cf. *nyssicola*	New Brunswick, CA	Blueberry
E-254	cf. *Nigrospora* sp.	New Brunswick, CA	Blueberry
E-255	cf. *Claussenomyces* sp.	New Brunswick, CA	Blueberry
E-256	*Phyllosticta* cf. *pyrolae*	Mt. Thom, NS	Blueberry
E-257	cf. *Mycosphaerella* sp.	New Brunswick, CA	Blueberry
E-258	*Godronia* cf. *cassandrae*	New Brunswick, CA	Blueberry
E-259	*Fusarium tricinctum*	New Brunswick, CA	Blueberry
E-260	*Sphaerulina* cf. *rhabdoclinis*	Portapique, NS	Blueberry
E-261	*Nigrospora sphaerica*	Simcoe, ON	Raspberry

E-262	*Colletotrichum fioriniae*	New Brunswick, CA	Blueberry
E-263	*Paraphaeosphaeria neglecta*	Mt. Thom, NS	Blueberry
E-300	cf. *Tumularia aquatica*	AAFC Kentville, NS	Blueberry
E-301	*Nemania serpens*	AAFC Kentville, NS	Blueberry
E-302	*Venturia hystrioides*	AAFC Kentville, NS	Blueberry
E-303	*Xylaria ellisii*	AAFC Kentville, NS	Blueberry
E-304	*Pyrenochaeta* cf. *cava*	AAFC Kentville, NS	Blueberry
E-305	*Pyrenochaeta* cf. *cava*	AAFC Kentville, NS	Blueberry
E-306	*Pseudoplectania nigrella*	AAFC Kentville, NS	Blueberry
E-307	PCR unsuccessful	AAFC Kentville, NS	Blueberry
E-308	*Aspergillus amstelodami*	AAFC Kentville, NS	Blueberry
E-309	PCR unsuccessful	AAFC Kentville, NS	Blueberry
E-310	*Cladosporium* cf. *cladosporioides*	AAFC Kentville, NS	Blueberry
E-312	*Pyrenochaeta* cf. *cava*	AAFC Kentville, NS	Blueberry
E-313	*Pyrenochaeta* cf. *cava*	AAFC Kentville, NS	Blueberry
E-314	*Pyrenochaeta* cf. *cava*	AAFC Kentville, NS	Blueberry
E-316	*Pseudoplectania episphagnum*	AAFC Kentville, NS	Blueberry
E-317	*Xylaria ellisii*	AAFC Kentville, NS	Blueberry
E-318	cf. *Leptodontidium* sp.	Mt. Thom, NS	Lowbush blueberry
E-319	*Paraphaeosphaeria neglecta*	Debert, NS	Lowbush blueberry
E-320	*Paraphaeosphaeria neglecta*	Debert, NS	Lowbush blueberry
E-321	*Pyrenochaeta* cf. *cava*	Debert, NS	Lowbush blueberry
E-322	*Paraphaeosphaeria neglecta*	Debert, NS	Lowbush blueberry
E-323	*Pyrenochaeta* cf. *cava*	AAFC Kentville, NS	Blueberry
E-324	*Proliferodiscus* sp.	AAFC Kentville, NS	Blueberry
E-326	*Proliferodiscus* sp.	AAFC Kentville, NS	Blueberry
E-327	*Penicillium* sp.	AAFC Kentville, NS	Blueberry
E-330	*Proliferodiscus* sp.	AAFC Kentville, NS	Blueberry
E-331	*Proliferodiscus* sp.	AAFC Kentville, NS	Blueberry
E-332	*Proliferodiscus* sp.	AAFC Kentville, NS	Blueberry
E-333	*Pyrenochaeta* cf. *cava*	AAFC Kentville, NS	Blueberry
E-334	*Gloeophyllum sepiarum*	Mt. Thom, NS	Lowbush blueberry
E-335	*Mycosphaerella tassiana*	Debert, NS	Lowbush blueberry
E-336	*Plagiostoma petiolophilum*	Mt. Thom, NS	Lowbush blueberry
E-337	*Cytospora ribis*	Debert, NS	Lowbush blueberry

E-338	*Pleonectria rosellinii*	Kemptown, NS	Lowbush blueberry
E-339	*Pyrenochaeta* cf. *cava*	Debert, NS	Lowbush blueberry
E-340	cf. *Leptodontidium* sp.	Debert, NS	Lowbush blueberry
E-341	cf. *Leptodontidium* sp.	Debert, NS	Lowbush blueberry
E-342	*Mollisia* cf. *melaleuca*	Kemptown, NS	Lowbush blueberry
E-343	*Lachnellula calyciformis*	Kemptown, NS	Lowbush blueberry
E-344	*Pseudoplectania* cf. *nigrella*	Debert, NS	Lowbush blueberry
E-345	*Peniophora* cf. *rufa*	Mt. Thom, NS	Lowbush blueberry
E-346	*Hypoxylon submonticulosum*	AAFC Kentville, NS	Blueberry
E-347	*Xylaria ellisii*	AAFC Kentville, NS	Blueberry
E-349	*Trichaptum* cf. *abietinum*	AAFC Kentville, NS	Blueberry
E-350	*Lachnellula calyciformis*	Kemptown, NS	Lowbush blueberry
E-351	*Sarea* cf. *difformis*	Mt. Thom, NS	Lowbush blueberry
E-352	*Discostroma* cf. *fuscellum*	AAFC Kentville, NS	Blueberry
E-353	*Proliferodiscus* sp.	AAFC Kentville, NS	Blueberry
E-354	*Xylaria ellisii*	AAFC Kentville, NS	Blueberry
E-356	*Pyrenochaeta* cf. *cava*	AAFC Kentville, NS	Blueberry
E-357	*Discostroma* cf. *fuscellum*	AAFC Kentville, NS	Blueberry
E-358	*Claviceps purpurea*	Mt. Thom, NS	Lowbush blueberry
E-360	*Botrytis cinerea*	Mt. Thom, NS	Lowbush blueberry
E-361	*Proliferodiscus* sp.	AAFC Kentville, NS	Blueberry
E-362	cf. *Tumularia aquatica*	AAFC Kentville, NS	Blueberry
E-363	*Proliferodiscus* sp.	AAFC Kentville, NS	Blueberry
E-366	*Pyrenochaeta* cf. *cava*	AAFC Kentville, NS	Blueberry
E-367	From Xylariales order	AAFC Kentville, NS	Blueberry
E-368	*Cytospora* cf. *cedri*	AAFC Kentville, NS	Blueberry
E-369	*Pyrenochaeta* cf. *cava*	AAFC Kentville, NS	Blueberry
E-370	*Alternaria* cf. *alternata*	AAFC Kentville, NS	Blueberry
E-372	*Discostroma* cf. *fuscellum*	AAFC Kentville, NS	Blueberry
E-373	*Paraphaeosphaeria neglecta*	Debert, NS	Lowbush blueberry
E-375	*Xylaria ellisii*	AAFC Kentville, NS	Blueberry

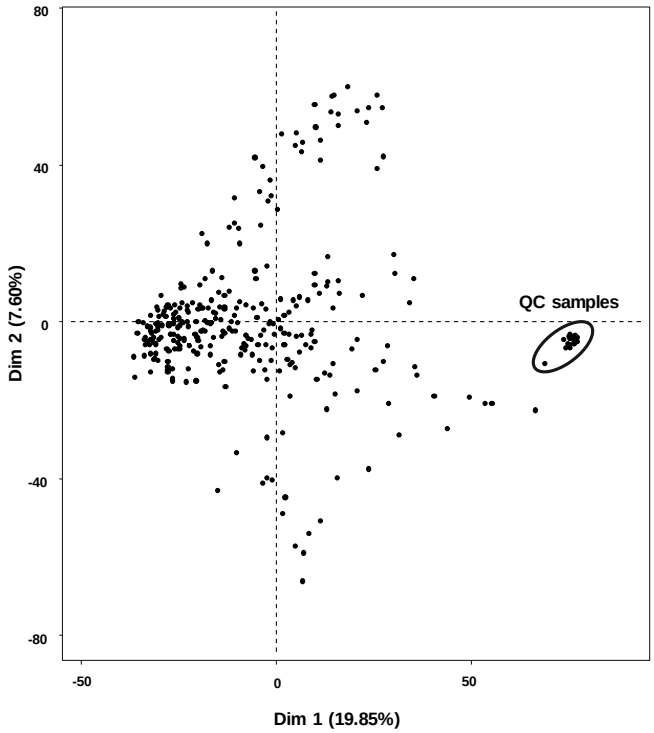


Figure C1. PCA plot of 302 endophyte extracts based on positive ionization mode HMRS data. Quality Control (QC) samples—circled in black—are all grouped tightly together, indicating little to no instrumental drift over the course of analysis.

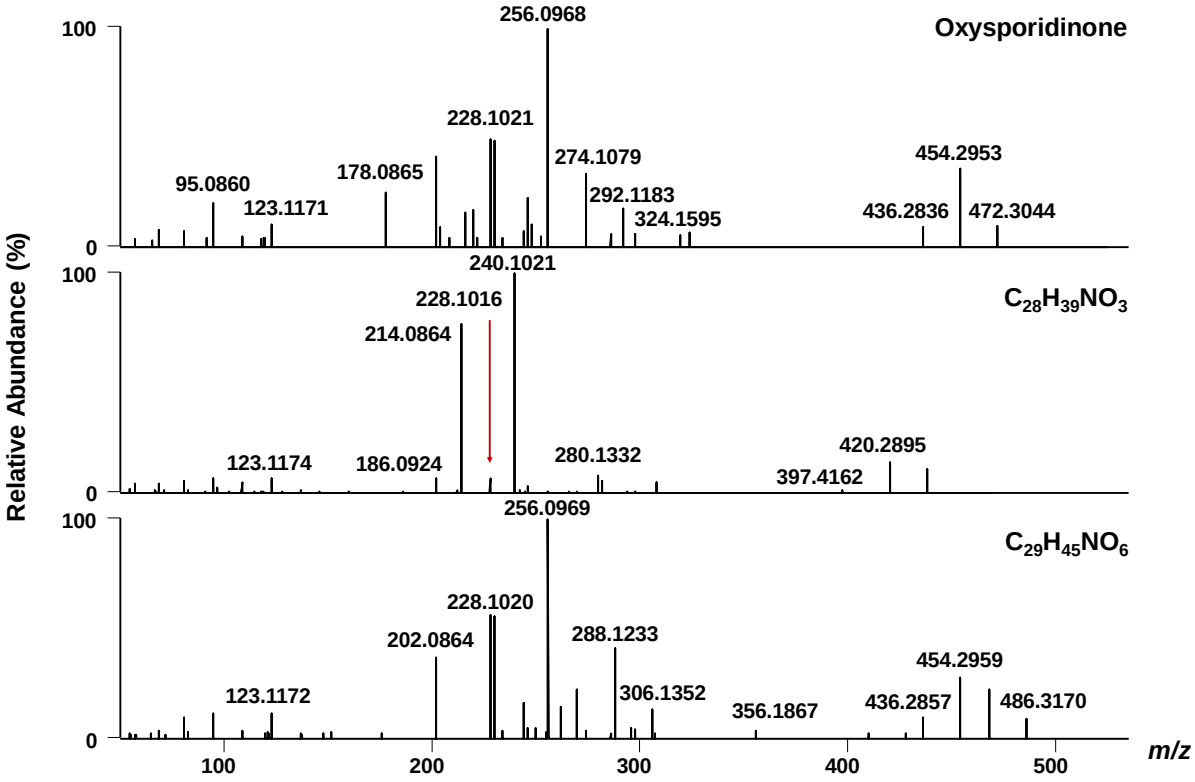


Figure C2. Tandem mass spectra of oxysporidinone and two unknown structurally related compounds associated within cluster F.

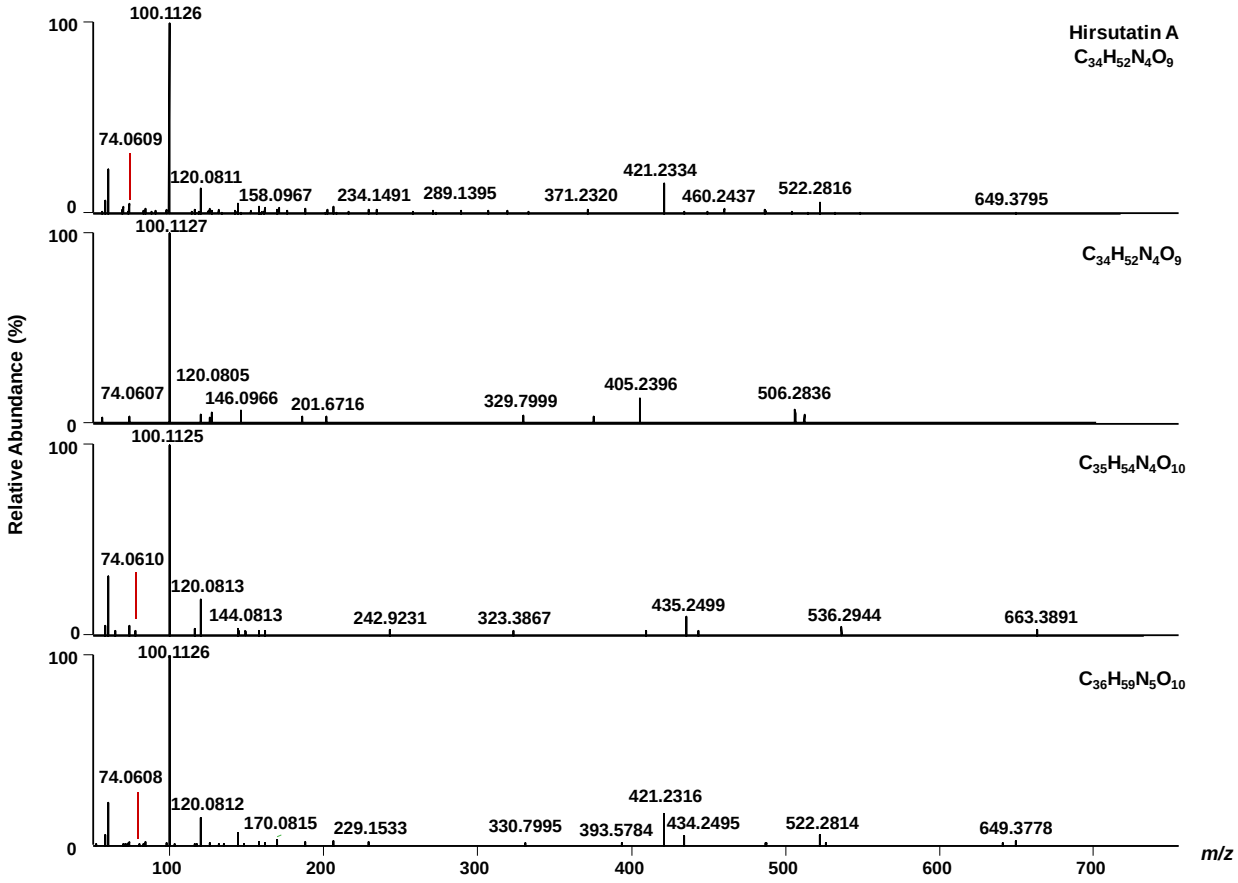


Figure C3. Tandem mass spectra of hirsutatin A and three unknown structurally related compounds associated within cluster R.

Natasha DesRochers

Curriculum Vitae

Natasha DesRochers, MSc Student
Agriculture and Agri-Food Canada (AAFC) and Western University

Post-Secondary Degrees

MSc, Analytical Chemistry **Sept 2018-Aug 2020**
Western University-London, ON
London Research and Development Centre, AAFC
Supervised by Dr. Mark Sumarah (AAFC) and Dr. Ken Yeung (Western University)

BMSc, Honours Specialization in Interdisciplinary Medical Sciences **2013- 2018**
Western University-London, ON

Honours and Awards

Dean's Honor List 2018

Dean's Honor List 2014

Western Scholarship of Excellence 2013

Work Experience

Teaching Assistant (Organic Chemistry) **Sept 2018 – Apr 2020**
Western University-London, ON

- Supervised laboratory sessions for four groups of 24 second-year students
- Gave weekly pre-lab instructional talks on lab technique and on theory behind experiments

Co-op Student **May 2016 – Aug 2018**
London Research and Development Centre, AAFC

- Performed liquid-liquid extractions using various solvents to acquire crude extracts
- Isolated fungal natural products by HPLC and flash chromatography
- Analyzed samples by HRMS using targeted and untargeted MS methods
- Analyzed secondary metabolite production of fungal species using
- Maintained microbiological cultures using meticulous aseptic technique
- Determined LOD and LOQ of analytes and created standard curves for quantification

Scientific Publications

DesRochers, N., Walsh, J. P., Renaud, J. B., Seifert, K. A., Yeung, K. K. C., & Sumarah, M. W. (2020). Metabolomic Profiling of Fungal Pathogens Responsible for Root Rot in American Ginseng. *Metabolites*, *10*(1), 35. DOI: 10.21228/M8KM4W

Walsh, J. P., **DesRochers, N.**, Renaud, J. B., Seifert, K. A., Yeung, K. K. C., & Sumarah, M. W. (2019). Identification of N, N′, N ″-triacetylfusarinine C as a key metabolite for root rot disease virulence in American ginseng. *Journal of Ginseng Research*. DOI: 10.1016/j.jgr.2019.08.008

Scientific Conferences

Ginseng Replant Disease Working Group Meeting **Jan 2020**

- Gave an oral presentation entitled “Metabolomic profiling of ginseng root pathogens”

Gordon Research Conference-Mycotoxins and Phycotoxins **Jun 2019**

Stonehill College-Easton, MA.

- Presented a poster on original research
- Attended talks on new findings in the field of mycotoxins

Great Lakes Mycology Meeting **May 2019**

Brock University-St. Catherines, ON

- Gave a brief oral presentation on original research

Post-ASMS Symposium **Oct 2018**

Crowne Plaza-Toronto, ON.

- Attended seminars on novel mass spectrometry methods and results

Summary of Coursework

Chemistry 9544A – Mass Spectrometry – 95

Chemistry 9522A – Analytical Separations – 90

Chemistry 9657Y – Graduate Seminar – Completed